Application of Biomarkers in Cancer Epidemiology

Edited by P. Toniolo, P. Boffetta, D.E.G. Shuker,
N. Rothman, B. Hulka and N. Pearce

IARC Scientific Publications No. 142

International Agency for Research on Cancer, Lyon, 1997

Published by the International Agency for Research on Cancer,
150 cours Albert Thomas, F-69372 Lyon cedex 08, France

Distributed by Oxford University Press, Walton Street, Oxford, UK OX2 6DP (Fax: +44 1865 267782) and in the USA by Oxford University Press, 2001 Evans Road, Carey, NC 27513, USA (Fax: +1 919 677 1303).
All IARC publications can also be ordered directly from IARC*Press*
(Fax: +33 4 72 73 83 02; E-mail: press@iarc.fr).

IARC Library Cataloguing in Publication Data

Application of biomarkers in cancer epidemiology/
editors, P. Toniolo ... [et al.]
(IARC scientific publication ; 142)

1. Genetic Markers
2. Neoplasms - epidemiology
3. Neoplasms - prevention & control I. Toniolo, P. II. Title III. Series

PDF ISBN 978-92-832-2211-8 (NLM Classification: W1)
ISSN 0300–5085

Foreword

Although biological markers have been used in epidemiology for decades, it is only in recent years, that advances in cellular and molecular biology have greatly expanded their potential. Biomarkers have possibilities in measuring whole-body or organ-specific exposures, as indicators of biological change or of early disease, and in assessing individual susceptibility to exposure. These developments foster the incorporation of a more biological thinking into studies of cancer epidemiology and open up new perspectives for elucidating the mechanisms of pathogenesis *in vivo* at the cellular and molecular level. Although the promise that biological markers hold in expanding the boundaries of epidemiological research is enormous, the criteria for their effective use in human observational research are as yet poorly understood, as are their limitations.

Biomarkers have many useful future applications in cancer prevention and public health; particularly in the monitoring of exposure to hazardous substances, in identifying individuals at increased risk of disease, and in risk assessment at the population level.

This volume aims to provide a set of state-of-the-art reviews of methodological issues in the use of biological markers in cancer epidemiology. In addition, it provides a brief expert statement of our present understanding of these issues under three broad headings: study design and analysis; development of biomarkers; and application of biomarkers.

The support of the United Kingdom Department of the Environment and the collaboration of the New York University are gratefully acknowledged. We thank Fred Garber for kindly designing the artwork for the cover of the book.

P. Kleihues
Director, IARC

International Agency For Research On Cancer

The International Agency for Research on Cancer (IARC) was established in 1965 by the World Health Assembly, as an independently financed organization within the framework of the World Health Organization. The headquarters of the Agency are in Lyon, France.

The Agency conducts a programme of research concentrating particularly on the epidemiology of cancer and the study of potential carcinogens in the human environment. Its field studies are supplemented by biological and chemical research carried out in the Agency's laboratories in Lyon and, through collaborative research agreements, in national research institutions in many countries. The Agency also conducts a programme for the education and training of personnel for cancer research.

The publications of the Agency contribute to the dissemination of authoritative information on different aspects of cancer research. A complete list is printed at the back of this book. Information about IARC publications, and how to order them, is also available via the Internet at: **http://www.iarc.fr/**

Workshop on the Application of Biomarkers to Cancer Epidemiology, Lyon, 20–23 February 1996
List of participants

R.J. Albertini
Genetics Laboratory, University of Vermont, 32 North Prospect St, Burlington, VT 05401-0508, USA

B.K. Armstrong
Cancer Control Information Centre, NSW Cancer Council P.O. Box 572, Kings Cross, NSW 2011, Australia

H. Bartsch
German Cancer Research Centre, Im Neuenheimer Feld 280 69120 Heidelberg, Germany

C.W. Boone
Chemoprevention Branch Division of Cancer Prevention and Control, National Cancer Institute, Bldg 31, Rm 10A52 31 Center Drive, Bethesda, MD 20892-2580, USA

F.X. Bosch
Servei d'Epidemiologia Hospital Duran i Reynals Autovia de Castelldefels, Km 2.7 08907 Hospitalet de Llobregat Barcelona, Spain

N. Caporaso
Genetic Epidemiology Branch National Cancer Institute, EPN 439, 6130 Executive Boulevard, Bethesda, MD 20892-7377, USA

D. Coggon
MRC Environmental Epidemiology Unit University of Southampton Southampton General Hospital Southampton SO16 6YD, UK

P. Farmer
MRC Toxicology Unit, University of Leicester, Hodgkin Building, Lancaster Road, P.O. Box 138, Leicester LE1 9HN, UK

S. Garte
Program in Environmental Carcinogenesis, NY University Medical Center, MSB Room 209, 550 First Avenue New York, NY 10010, USA

D. Gompertz
MRC Institute for Environment and Health, University of Leicester, 94 Regent Street Leicester LE1 9HN, UK

K. Hemminki
Center for Nutrition and Toxicology, Karolinka Institute, Hälsovagen 7, 14157 Huddinge, Sweden

B. Hulka
Department of Epidemiology School of Public Health University of North Carolina CB No. 7400 McGavran-Greenberg Hall, Chapel Hill, NC 27599-7400, USA

D.J. Hunter
Channing Laboratory, Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115, USA

M.T. Landi
Genetic Epidemiology Branch National Cancer Institute Executive Plaza North, Rm 439, 6130 Executive Boulevard, Msc 7372, Bethesda, MD 20892-7372, USA

A.J. McMichael
Department of Epidemiology London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

N. Pearce
Wellington Asthma Research Group, Department of Medicine, Wellington School of Medicine, P.O. Box 7343 Wellington South, New Zealand

F. Perera
Division of Environmental Sciences, Columbia University School of Public Health Room B109, 60 Haven Avenue New York, NY-10032, USA

J.D. Potter
Head, Cancer Prevention Research Program Fred Hutchinson Cancer Research Center 1124 Columbia Street Seattle, WA 98104, USA

N. Rothman
Environmental Epidemiology Branch, National Cancer Institute, EPN 418, 6130 Executive Boulevard, Bethesda, MD 20892-7364, USA

R. Saracci
Istituto Fisiologia Clinica CNR Epidemiology Section Via Trieste, 41 56100 Pisa, Italy

A. Schatzkin
Division of Cancer Prevention Studies Branch, National Cancer Institute, 6130 Executive Boulevard, Bethesda, MD 20852, USA

P.A. Schulte
Screening and Notification Section National Institute for Occupational Safety and Health 4676 Columbia Parkway, Room 42, Cincinnati, OH 45226-1998, USA

D. Shuker
Toxicology Unit University of Leicester Hodgkin Building, Lancaster Road, P.O. Box 138 Leicester LE1 9HN, UK

E. Taioli
Department of Environmental Medicine, NYU School of Medicine, 341 East 25th Street, 2nd Floor, New York, NY 10010, USA

P. Toniolo
Head, Unit of Cancer Epidemiology, New York University School of Medicine, 341 East 25th St, New York, NY 10010-2598, USA

E. White
Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center 1124 Columbia Street, MP-702 Seattle, WA 98104, USA

Z.F. Zhang
Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center 1275 York Avenue, B ox 44 New York, NY 10021, USA

Observers

H. Autrup
Department of Environmental and Occupational Medicine Aarhus University, Universitetsparken bygning 180, 8000 Aarhus C., Denmark

T. Ballard
Venetian Tumour Registry Via Giustiniani, 2, 35100 Padua, Italy

IARC participants

P. Boffetta
M. Freisen
P. Hainaut
J. Hall
R. Montesano
N. Muñoz
P. Pisani
E. Riboli
C. Wild

Contents

Application of Biomarkers in Cancer Epidemiology

(IARC Scientific Publications No. 142)

Errata

- List of participants attending the Workshop on the Application of Biomarkers to Cancer Epidemiology, Lyon, 20-23 February 1996, page v, should include the following participants:

  R. Sinha
  Nutritional Epidemiology Branch, National Cancer Institute, EPN 443, Bethesda, MD 20892, USA

  J.D. Tucker
  Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, P.O. Box 808, L-452, Livermore, CA 94551-9900, USA

- The Technical Editor for the Workshop, page v, was:
  S. Garber
  Epidemiology Program, New York Univ. School of Medicine, 341 East 25th St, New York, NY 10010-2598, USA

- The following persons contributed to the Workshop proceedings, page v, but were unable to attend:

  P. Vineis
  Cancer Epidemiology Unit, University of Turin, Via Santena, 7, I-10126 Turin, Italy

  R. Kaaks
  NTR/IARC

- In line 12 of the first paragraph of p. 309, replace 95.7% with 9.75%.

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Application of biomarkers in cancer epidemiology

Workshop report*

In epidemiology, a biological marker (commonly abbreviated, for convenience, to biomarker) is any substance, structure or process that can be measured in the human body or its products and may influence or predict the incidence or outcome of disease. Biomarkers can be broadly classified into markers of exposure, effect and susceptibility. Biomarkers may include the following: xenobiotic agents and their metabolites in tissues or body products; normally occurring body constituents whether in physiological or pathological amounts; endogenous compounds that are not present under normal conditions; and inherited and acquired abnormalities of body chemistry, structure or function, including pathological manifestations of precursors to disease. Biomarkers should be distinguished from biomarker assays, specific laboratory tests aimed at measuring particular biomarkers, and biomarker measurements, the amounts of particular biomarkers present in specified units of tissues or body products as measured by biomarker assays.

Biomarkers have been used in epidemiological studies of cancer for many years. Early examples include the classic studies of B. MacMahon and colleagues on the geographical correlation of urinary estrogen concentrations with cancer of the breast (MacMahon *et al.*, 1974; Dickinson *et al.*, 1974), subsequent studies of the relationship between urinary and blood estrogens and breast cancer in case–control studies (Cole & McMahon, 1969; MacMahon *et al.*, 1982, 1983) and analyses of the relationship of cancer mortality to serum cholesterol concentrations in the Framingham and other cohort studies (McMichael *et al.*, 1984). Two recent, major contributions of biomarkers to epidemiological studies are the demonstration of the carcinogenicity of aflatoxins in combination with hepatitis B in humans through a cohort study of liver cancer, including measurements of urinary metabolites and nucleic acid adducts of aflatoxin (Ross *et al.*, 1992; Qian *et al.*, 1994), and the identification of the major role of human papillomaviruses (HPVs) in causing cancer of the cervix worldwide (Muñoz *et al.*, 1992; Bosch *et al.*, 1995).

In spite of the quite extensive current use of biomarkers in cancer epidemiology, methodological aspects of their use have not been extensively elucidated or described. While there have been important contributions to this subject, including textbooks (Hulka *et al.*, 1990; Schulte & Perera, 1993) and comprehensive meeting reports (Mendelsohn *et al.*, 1995; Institute for Environment and Health, 1996) on the use of biomarkers in epidemiology, there remains a substantial need for relevant methodological research, wider discussion and understanding of the methodological issues, and their incorporation into the formal and informal training of cancer epidemiologists and other scientists pursuing careers in biological research on cancer.

The general objective of the use of biomarkers in cancer epidemiology is the same as that of cancer epidemiology itself: to gain knowledge about the distribution and determinants of disease occurrence and outcome that may be applied to reduce the frequency and impact of disease in human populations. There are, however, several specific objectives for the use of biomarkers in cancer epidemiology which should guide the evaluation of proposals to develop or apply biomarkers. Principally, they are to increase the accuracy of measurements of genetic or other acquired susceptibility to disease; of exposures that may cause or prevent disease; of exposures that confound or modify the associations between risk and other exposures; of disease itself; and of factors that may determine the outcome of disease, such as disease

*This paper is the consensus report of the workshop *Application of Biomarkers in Cancer Epidemiology*, which was held at the International Agency for Research on Cancer, Lyon, France, in February 1996.

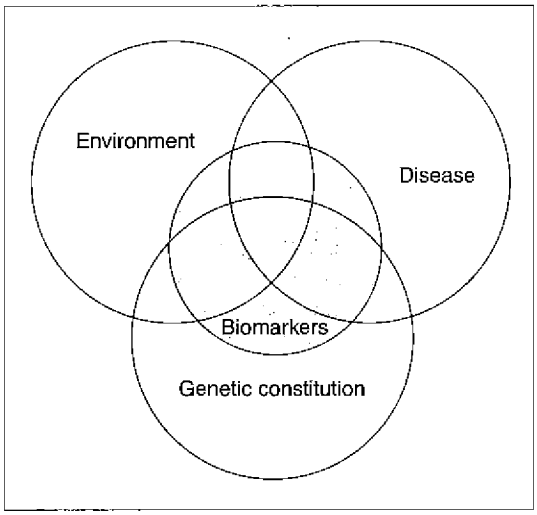

Figure 1. Schematic representation showing how biomarkers may be used to measure phenomena that reflect the amount or effects of agents that influence or predict disease incidence or outcome.

precursors and stages. Biomarkers may also be used to reduce, in the proposed study, the time interval between the relevant exposure and measurement of the putative effect and to increase the yield of information on disease pathogenesis. Ultimately, they should serve to increase the cost-effectiveness of epidemiological studies, in the sense that, as a result of their use, more information is gained per unit cost.

Because of their origin in the structure and function of the human body, biomarkers may provide more complex measures of the underlying exposure, process, etc. that they represent than other measures that are common in epidemiology. This complexity is illustrated in Fig. 1. The three outer circles, *genetic factors*, *environment* (broadly defined) and *disease*, represent the universe of phenomena that can be measured to reflect the amount or effects of exogenous or endogenous agents that may influence or predict the incidence or outcome of disease. The inner circle, *biomarkers*, represents the subset of these phenomena that can be measured in the human body. That almost all of the intersections between the three outer circles fall within the inner circle illustrates pictorially that biomarkers, more than any other epidemiological measure, may measure elements of any two, or all three, of genetic factors, environment and disease. For example, the activity of an inducible, carcinogen-metabolizing enzyme is influenced by both exposure to inducing agents and the host genetic constitution; blood cholesterol concentration, an indirect indicator of fat nutrition, is almost certainly influenced by the presence of early cancer (International Collaborative Group, 1982; Sherwin *et al.*, 1987); presence of CC→TT mutations in the *p53* gene in normal skin, a plausible marker of lifetime exposure to UV radiation (Nakazawa *et al.*, 1994), is also influenced by DNA repair enzymes (genetic susceptibility) and may be influenced by the pathological consequences of *p53* gene mutation (e.g. a proliferative advantage for the mutated cells). The practical consequences of this complexity have not yet been elucidated.

In many cases, as measurements get closer to the biologically effective dose of an external agent at its target tissue, the relationship between the biomarker and disease will be less likely to reflect the relationship between the amount of *external* exposure to the agent and disease. However, exposures from multiple sources, such as nitrate, which may include exogenous pathways, may be better evaluated using biomarkers.

Because of the complexity of the interrelationship between biomarkers, their potential for increasing knowledge about cancer epidemiology and pathogenesis is now emerging. Research is needed to explore the methodological characteristics and biological functions of specific biomarkers and how they relate to one another. Most particularly, their application in epidemiological studies should be undertaken after thorough preparation and interdisciplinary collaboration. Even then, interpretation of results must be undertaken with appropriate caution.

Epidemiological study design and analysis

The use of biomarkers of exposure or dose, disease or individual susceptibility aims to contribute to the elucidation of the causal relationships in human populations between diseases and factors such as external exposures (via personal habits, occupation and the ambient environment), genetic traits and interventions.

The types of epidemiological studies used in biomarker research closely parallel the types used in other fields of epidemiology; biomarker studies,

however, raise specific methodological issues which are addressed in the following sections.

Transitional studies

Transitional studies provide a bridge between the use of biomarkers in laboratory experiments and their use in cancer epidemiology studies (Schulte & Perera, this volume). They have as their outcome the characterization of biomarkers and of the problems in their use; thus, they yield preliminary results rather than end results about cancer etiology and prevention. Their intent should be to lay the groundwork for the use of a marker in full-scale epidemiological studies by addressing the following aspects: intra- and intersubject variability; the feasibility of marker use in field conditions; potential confounding and effect-modifying factors for the marker; and mechanisms reflected by the biomarker. Transitional studies can be divided into three functional categories: developmental, characterization and applied studies.

Developmental studies involve determining the biological relevance, pharmacokinetics, reproducibility of measurement of the marker, and the optimal conditions for collecting, processing and storing biological specimens in which the marker is to be measured.

Characterization studies involve assessing interindividual variation and the genetic and acquired factors that influence the variation of biomarkers in populations. This includes assessing the frequency or level of a marker in populations, identifying factors that are potential confounders or effect modifiers, and establishing the components of variance in the biomarker measurement, laboratory variability, intra-individual variation and interindividual variation. The ratio of intra-individual variation to interindividual variation has important implications for study size and power.

Applied transitional studies assess the relationship between a marker and the event that it marks, namely exposure, pre-clinical effects, disease or susceptibility. Applied studies are often of cross-sectional or short-term longitudinal design, and are not intended to establish or refute a causal relationship between a given exposure and disease. Rather, they are intended to assess whether relevant correlations exist and if they are strong enough to be useful in full epidemiological studies.

Like other studies using biomarkers, transitional studies raise ethical issues when the meaning of the biomarker results is not known. The objectives of the research generally are not to identify health risks but to identify characteristics of the biomarker or distribution by population subtypes. There is a need to anticipate the impact of transitional studies on study participants and plan to address their concerns (Hunter, this volume).

Cohort and case–control studies

Cohort and case–control studies represent the two most widely used types of observational study (Potter, this volume). They can be population-based or based on patients at a hospital, clinic or practice. For simplicity, studies based on patients will be referred to throughout as clinic-based studies.

In clinic-based cohort studies, of treated patients or screened populations, the inclusion of biological measures of exposure and susceptibility is both methodologically sound and logistically feasible. In population-based studies, collection of biological material for such markers is feasible but logistically more complex. For early outcome markers, collection of material (e.g. precancerous lesions) is logistically feasible in a hospital setting, but becomes more difficult in the population setting. End-points assessed routinely at a variety of institutes may produce major problems of standardization of methods (e.g. histological diagnosis). Except for some routinely registered biological characteristics (e.g. receptor status in breast cancer), even greater problems attend the issue of identification of biologically or pathologically defined disease subsets.

Most of the considerations that make prospective cohort studies an attractive study design apply to all methods of exposure assessment, including biomarkers. The strengths of such studies include the fact that exposure is measured before the outcome, that the source population that gave rise to the cases is explicitly defined, and that participation can be as high as 100% if specimens are available for all subjects and follow-up is complete. Weaknesses of prospective cohort studies involving biomarkers include the usually small number of cases of each of many types of cancer, the lack of specimen if the biomarker requires large amounts of specimen or unusual specimens, degradation of the biomarkers during long-term storage,

and the lack of detail on other potentially confounding or interacting exposures. Cohort studies are, in principle, the preferred method for studying the temporal sequence of intermediate end-points; however, the costs and complexity of repeated screening may often make this difficult.

Important issues in prospective cohort studies involving biomarkers, especially of exposure, involve the frequency and timing of specimen collections. A major concern in cohort studies of short duration (as in case–control studies) is the possibility that the disease process has influenced the biomarker level among cases diagnosed within 1–2 years of the specimen being collected. In studies of longer duration, there may be considerable misclassification of the aetiologically relevant exposures if specimens have been collected only at baseline. This misclassification occurs not only because an individual's exposure level may change systematically over time, but also because there may be considerable intra-individual variation (from day to day or even hour to hour) in biomarker levels. The effects of intra-individual variation can be reduced by taking multiple samples, but this greatly increases the expense of sample collection and storage and the burden on study subjects; similar considerations apply to taking samples at several points in time in an attempt to estimate time-integrated exposures or exposure change. An alternative approach is to estimate the extent of intra-individual variation, and the misclassification involved in taking single specimens, by taking multiple specimens in a sample of the cohort. This information can be used to correct for bias to the null introduced if the misclassification is non-differential, and therefore de-attenuate observed relative risks.

Use of biomarkers of disease susceptibility in cohort studies raises significant ethical issues, particularly if there is repeated contact with study subjects. In particular, informing the cohort members of their biomarker level is problematic if the biomarker is not considered to be sufficiently predictive of disease, and if there are no known preventive steps cohort members can take to reduce their risk of the disease.

A nested case–control study or a case-cohort design can be used to reduce the expenses of data collection and sample analysis. In a nested case–control study, biological samples may be collected from all cohort members and stored at the beginning of the study; the controls are then selected throughout the course of the study, ideally at the time that each case is diagnosed. An advantage of the nested case–control approach is that biomarkers can be measured in specimens matched on storage duration and case–control sets can be analysed in the same laboratory batch, reducing the potential for bias introduced by sample degradation and laboratory drift. On the other hand, a case–cohort design may be used when sample collection from the entire cohort is not feasible, or when the cost of storage or analysis of each sample is prohibitive. This design involves collecting biological samples from the cohort at the beginning of follow-up and then collecting samples from cases as they occur. However, as samples are being taken at different times for cases and 'controls' (i.e. the reference sample from the entire cohort), bias will be introduced if sample degradation or laboratory drift occurs over time. Furthermore, the case–cohort approach may lead to laboratory personnel being unblinded to case and control status.

In case–control studies, biomarkers of internal dose or effective dose are appropriate when they are stable over a long period of time (e.g. carrier status for infectious agents, such as HBsAg) or when exposures have been constant over the relevant exposure period; however, it is essential that they are not affected by the disease process, diagnosis or treatment (except in those circumstances when it is possible to collect specimens prior to the commencement of treatment). For genetic susceptibility markers, case–control studies are highly appropriate. Clinic-based case–control studies are particularly suitable for studies of intermediate end-points, as these end-points can be systematically measured. Population surveys (or screening) are a prerequisite for the identification of cases in population-based case–control studies of intermediate end-points.

Clinic-based case–control studies are of particular value for studying the etiology of precancerous lesions (e.g. early cervical intraepithelial neoplasia). None the less, the relation between these intermediate disease markers and the cancers of which they are precursors is usually difficult to study for both ethical and logistical reasons. In case–control studies involving cancer as the end-

point, biomarkers of exposure (though not of any given genotype) are of limited value, as the disease itself, its therapy, or behaviour changes in patients, may alter any exposure marker or biological process thought to be a precursor of disease.

Case–case comparisons can be useful in the study of biologically defined disease subsets when differences in the etiology of such subsets are being examined. However, if a control group is not included, any heterogeneity in the exposure associations among disease subsets may be the result of increased risk in one subset as opposed to a decreased risk in another.

Intervention studies

Intervention studies (trials) assess prospectively whether an intervention is efficacious in changing the frequency of a marker of exposure or dose (e.g. measures to encourage smoking cessation and either serum cotinine or protein or DNA adducts) or pre-clinical condition or disease (e.g. *p53* mutation, dysplasia and cell proliferation in trials using antioxidants to prevent second tumours). Biomarkers can be particularly valuable in assessing compliance with the intervention (e.g. serum β-carotene in a randomized trial of β-carotene). Markers of susceptibility (e.g. *GSTM1* polymorphism) could be used to determine whether groups in various intervention arms are comparable. Intervention studies may also be conducted as transitional studies when there are still questions about the response of biomarkers following an intervention. Critical in all types of intervention studies is the need to address those same issues that are important in cohort studies, including laboratory drift, loss to follow-up, and intervening exposures that may affect the biomarker.

Ecological studies

Ecological studies use groups rather than individuals as units of observation. They include studies of geographical differences and time trends in disease incidence and prevalence. While ecological studies provide useful information on exposure, disease and modifying factors, they are, in most cases, inadequate to establish causal relationships. Although, in general, ecological studies based on biomarkers have the same advantages and limitations as the other types of ecological studies, it is possible to conduct carefully planned ecological studies, in particular in the control of confounding. Such studies often do not seek to make causal inferences about exposure– disease relationships; rather, they are aimed to elucidate plausible mechanisms (Kuopio *et al.*, 1995).

Cross-sectional studies

Cross-sectional study designs, with single or repeated sampling, have been frequently used in the validation of markers of exposure and of disease. They can provide a useful snapshot of the relationship between exposure and susceptibility factors, on the one hand, and the selected biomarker, on the other, in populations whose exposure can be well characterized. Frequently, the studies have involved industrial populations, chemotherapy patients and smokers—all groups exposed to 'model' carcinogens—and have compared them to appropriate non-exposed groups. Such studies can also be used in healthy individuals to establish the possible range of measurements of protective factors, e.g. monitoring of vegetable intake by measurement of carotenoids or phenolic compounds in plasma and urine. This design can be used to establish an association between the biomarker and the exposure under study.

Family-based studies

Family studies are intended to address scientific questions related to four general areas of research:

1. Does familial aggregation exist for a specific disease or characteristic?
2. Is the aggregation due to genetic factors or environmental factors, or both?
3. If a genetic component exists, how many genes are involved and what is their mode of inheritance?
4. What is the physical location of these genes and what is their function?

Using linkage methods, biomarkers that characterize known DNA haplotypes have been instrumental in locating a number of cancer-related genes. These studies require relatively large amounts of germline DNA (usually represented by DNA extracted from white blood cells). This is because there are multiple marker loci that require evaluation. Once a locus is identified, further DNA is necessary to identify the gene. Often, fibroblast or

Epstein–Barr virus (EBV)-transformed cell lines are required to provide adequate DNA for these studies.

Special concern must be paid to ethical and counselling issues in family studies in which biomarker results for an individual may also have predictive value for other members of the family.

General issues in study design and analysis

As with other epidemiological studies, the design and statistical analysis of a study involving biomarkers involves in general: (1) relating a particular disease (or marker of early effect); (2) to a particular exposure; while (3) minimizing bias; (4) controlling confounding; (5) assessing and minimizing random error; and (6) assessing interactions (Pearce & Boffetta, this volume).

Many studies using biomarkers of disease are of cross-sectional design and measure the prevalence of the disease state, which is dependent on both its incidence and its duration. Thus, in a study examining markers of cell damage as an effect of exposure to known or suspected carcinogens, the results would depend on factors such as the turnover of the cells in which the marker is measured and the capacity to repair the damage.

When deciding whether to measure exposure with biomarkers, it is important to consider how useful and informative the biomarkers are in relation to the study hypothesis. This is a particular issue for exposures that change over time; for example, a biomarker may be more valid than a questionnaire in assessing exposure to HPV, whereas a questionnaire will be more valid than a biomarker in assessing cigarette smoking 10 years previously. It should be stressed that the concept of 'useful and informative' depends on: (1) whether, in principle, it is most appropriate to study external exposure or internal dose (i.e. which hypothesis is being tested); (2) what information is available for the etiologically relevant time period; (3) whether the exposure or dose measurement is subject to intra-individual variation; and (4) whether it is desirable and possible to obtain (with a biomarker) information on interindividual variation in exposure or biological response.

When exposure or disease is measured with biomarkers, it is important to ensure that any misclassification is non-differential (i.e. it applies equally to the groups being compared) and is as small as possible. However, because the relationships between exposure, marker and disease are, in most cases, obscure, the capacity to establish the presence or absence of misclassification in the interpretation of the findings of biomarker-based studies is limited.

The use of biomarkers of exposure does not reduce the need to control for confounding and, in some instances, the use of biomarkers may actually introduce confounding into a study (Pearce & Boffetta, this volume). Confounding can be controlled in the analysis by stratifying the data into subgroups according to the levels of the confounder(s), or by the use of mathematical modelling. However, problems of multicollinearity can occur when variables that are highly correlated (e.g. serum levels of various micronutrients) are entered simultaneously into a model; this will make the model unstable and can lead to invalid effect estimates.

Some biomarkers may be surrogates for intermediate stages in the disease process. For example, when the relationship between reproductive risk factors and cervical cancer is adjusted for HPV status, the relative risks for number of sexual partners are reduced. This suggests that HPV is a mediator of the relationship between the reproductive risk factor and cervical cancer, helping to substantiate the causal role for HPV. In cases such as this, the biomarkers should not be treated as potential confounders, i.e. they should not be adjusted for in the analysis. None the less, analyses alternatively including and excluding the putative intermediate can be used to help determine causal pathways.

When choosing the method of exposure or dose assessment, it is important to consider their implications on study size, particularly when an expensive or invasive method may be used. An additional consideration in study size estimation is the ratio of the number of assays per individual and the number of individuals in the study. Many biomarkers show marked variation from day to day within the same individual (in part from problems of repeatability of the laboratory tests, but also from genuine differences), and in some cases the intra-individual variation may be greater than the interindividual variation (e.g. 24-h urinary sodium measurements). Thus, it is important to take into account the trade-off between including more participants, on the one hand, and gathering multiple samples from each participant, on the other. It is essential that

the relative magnitude of intra-individual variation of a biomarker is adequately characterized.

Biomarkers present better opportunities for assessing interactions between genetic and environmental factors. In particular, to infer interaction, biomarkers of genetic susceptibility should show a higher disease risk in exposed susceptible groups than in exposed non-susceptible and in non-exposed groups. None the less, such testing for interaction usually requires a substantial increase in study size.

As is the case for other aspects of epidemiological research, if the relative risk between a biomarker and exposure or outcome is not very strong (e.g. less than 10), then results from multiple studies will be required before agreement on the existence of an association can be reached. It is critical, therefore, that the results from all existing studies are available. Unfortunately, the tendency for investigators and journals to publish only 'positive' results may bias the literature in favor of positive studies, a phenomenon often called 'report bias' or 'publication bias'. Biomarker research may be even more prone to this bias than other fields of epidemiological research, as multiple biomarkers may be assessed, sometimes at relatively low cost. It is important that investigators publish, or make available in abstract form, all their results from studies in which the results are reliable, even if these results are 'null'. Publication bias can be particularly relevant when several analyses of the data are done on subgroups of the study population, defined according to characteristics such as tumour markers or susceptibility gene polymorphisms. Such multiple analyses increase the probability to obtain significant results by chance only. Although these analyses may provide useful information on effect modifiers, the selective reporting of significant or 'positive' results should be discouraged.

Incorporation of biomarkers into epidemiological studies

Biological markers undergo a process of discovery, characterization and refinement before use in epidemiological studies. Animal studies contribute to the development and validation of biomarkers and provide insights into the mechanisms of the multistage process of carcinogenesis. Transitional studies assist in this process by optimizing sample processing, evaluating intra-individual and inter-individual variation in the biomarker in target populations, determining laboratory assay variation, and studying the impact of exogenous and endogenous influences on the biomarker.

Regardless of the specific protocol used for the collection, processing and analysis of a specific biomarker, two important issues to consider are the storage of biological samples and the evaluation of sources of variation in biomarker measurement. This section will therefore review storage and variation issues following the discussion of individual biomarkers.

The following biomarkers have been selected to demonstrate the wide range of assays available for incorporation into epidemiological studies, and to comment on their utility for various study designs relevant to the study of cancer. This is not meant to be a comprehensive list; other important categories of biomarkers in blood (e.g. immunological markers) and in normal and preneoplastic tissue are not addressed here.

Biomarkers of biological agents

Biological agents associated with chronic infection and subsequent development of cancer are measured using serological or nucleic acid markers (Muñoz & Bosch, this volume). An example of nucleic acid-based biomarker is HPV DNA detection where the presence of type-specific DNA at a given time is measured by PCR-based assays. Cohort studies having high-grade squamous intraepithelial lesions as the end-point have shown that the infection preceded the disease. HPV DNA infections are often transient, especially in young women. Therefore, repeated sampling is required to assess persistent HPV infections. Accurate serological assays aimed at distinguishing transient from persistent infections need to be developed. Another example is the hepatitis B virus (HBV), for which there are serological markers that distinguish between past and persistent infections. HBV DNA detection in sera further refines the assessment of exposure. These markers have been used in all types of epidemiological design (cross-sectional, case–control, cohort and intervention studies).

Biomarkers of internal dose for chemical exposures

Biomarkers of internal dose of external chemical exposures are measurements of a parent compound

or its metabolite(s) in an accessible biological matrix, such as serum or urine (Institute for Environment and Health, 1996). They have potential applications in several types of epidemiological studies (Coggon & Friesen, this volume). These include studies to establish the importance of different sources of exposure as determinants of total dose; studies to validate other methods of exposure assessment such as the use of questionnaires; studies to establish that a chemical reaches a suspected target tissue; cross-sectional studies relating exposure/dose to biomarkers further downstream in the exposure–disease continuum; and case–control and cohort studies relating exposure/dose to disease.

The utility of an internal dose marker in case–control and prospective cohort studies depends, in part, upon the half-life of the external agent or its metabolites in the body; the pattern of the exposure it is measuring (e.g. regular, daily exposure versus infrequent, episodic exposure); whether secular trends have occurred in that exposure (e.g. smoking cessation); and direct or indirect influences of the disease process. The information that a biomarker of internal dose provides must be compared to the availability and quality of other sources of data (e.g. questionnaires, environmental measurements, medical records). Essentially, all exposure measures misclassify some subjects on their usual pattern of exposure; it is the relative ability of different sources of data to place individuals correctly into exposure categories that is important. In case–control studies, questionnaires remain the primary source of exposure data. There are, however, a few instances where a measure of internal dose may be a more suitable measure of cumulative exposure, particularly in the study of cases with preneoplastic lesions or early disease. The ideal biomarker should persist over time (e.g. fat-soluble substances such as DDT metabolites) and should not be affected by disease status. Markers of internal dose may be useful in prospective cohort studies as long as components of variance of the biomarker are well characterized, since the problem of reverse causality (i.e. disease status affects the level of the biomarker) is minimized.

Biomarkers of dietary intake and nutritional status

Dietary intake is usually assessed by various types of questionnaire and other methods (e.g. diaries); however, there are many sources of error involved in using this approach. Biomarkers can be useful in nutritional epidemiological studies at two different levels (Kaaks *et al.*, this volume). First, biomarkers may be of interest as potentially more precise, more specific or more objective measurements (or correlates) of the intake levels of specific foods or food constituents, compared to measurements obtained with questionnaire and interview methods. Secondly, most biomarkers can also be seen as indicators of a nutritional/metabolic status, which may be intermediate between 'exogenous' nutritional lifestyle factors (e.g. the composition of diet, physical activity) and disease risk. When used as an indicator of intake, they are useful in four different types of study: validation (comparison of a measurement against a gold standard), calibration (comparison of two measurements), observational epidemiological and intervention studies. They are especially important in validation or calibration studies where biomarkers represent an additional category of measurement, the 'random' errors (i.e. variations that are independent of individuals' true habitual intake levels) of which can be assumed to be statistically independent of those of the questionnaire and other methods. The factors that are important in deciding whether to use a given biomarker in a particular study include whether it is a good indicator of intake; whether it is a long- or short-term marker; whether there is need for multiple measurements; whether it is acceptable to the researcher and the subject; and whether it is compatible with the design of the study itself (case–control of early disease, case–control of late disease, or a cohort study). Biomarkers can also be informative if they provide an integrated biological measure of intake, lifestyle and metabolic processes. For example, measurements of red cell folate coupled with serum homocysteine would offer a picture of medium-term intake of folic acid, which is distinguishable from transient fluctuations in dietary intake (Green & Jacobsen, 1995).

Biomarkers for endogenous hormones

Biomarkers are available to measure specific endogenous hormones (Lemaster & Schulte, 1993; Hulka *et al.*, 1994). Cross-sectional studies, where the biomarker is the outcome measure, are important for the assessment of differences in marker levels between subjects with different characteristics (e.g. sex, race, anthropometry, geographical

location). A case–control study is not an appropriate study design to assess associations between metabolic and hormone markers and disease outcome, since marker sampling occurs after the clinical appearance of disease. This problem may be less significant or insignificant if the outcome is a pre-clinical condition or a very early lesion. Cohort studies, including case–control studies nested within cohort studies, are ideally suited for evaluating the association between hormone biomarkers and disease. Sampling issues important in cohort studies are the frequency of sampling and the timing of collection in relation to events which may influence measurement (e.g. stage of menstrual cycle, menopause, oophorectomy, ageing, medications).

Macromolecular adducts as biomarkers of exposure to reactive chemicals

Chemicals can bind covalently to cellular macromolecules such as nucleic acids and proteins (Wild & Pisani, this volume). The product of this addition of a chemical moiety to a macromolecule is termed an 'adduct'. The adduct may be highly specific for the carcinogen of interest, but not necessarily specific for a given exposure because of multiple sources of the carcinogen within the environment. Adduct formation normally occurs after the metabolic activation of the carcinogen; DNA repair may follow adduct formation. As a result, measured adducts represent an integration over time of carcinogen exposure and interindividual variations in carcinogen metabolism, DNA repair and other host factors. The persistence of adducts is determined by the chemical stability of the adduct itself and the turnover of the macromolecule to which the chemical is bound. In practice, this gives a half-life of adducts on proteins (haemoglobin and albumin) of a few weeks to months, while DNA adducts may have half-lives of a few hours to several years depending on the cell type concerned. Adducts of more remote exposure (such as modified amino acids in histone proteins from non-dividing cells) would represent a major advance in the utilization of these markers in epidemiological studies. Adduct measurements can be made in blood and exfoliated cells, and metabolites of adducts can be measured in urine. The quantity of material required is dependent on the assay sensitivity for a given adduct.

Cross-sectional studies evaluate exposure–adduct relationships in populations currently exposed to agents of concern. Elevated adduct levels in an 'exposed' versus 'unexposed' population may suggest that the exposure is associated with a higher cancer risk, given the increasing evidence that DNA damage represents a primary mechanism of carcinogenesis. In this instance, the adduct is not being used as a dosimeter, but rather as evidence of a potentially harmful response *in vivo*. Such studies should be regarded as providing supporting evidence only, until the association of adducts with subsequent development of cancer has been demonstrated. It is relevant to note, however, that in the presence of limited epidemiological evidence of a chemical's carcinogenicity, the demonstration in humans that the chemical causes a dose-dependent increase in macromolecular adducts or in other biomarkers that reflect genotoxic damage provides supporting evidence that the chemical is carcinogenic to humans. This type of evidence has been used, for example, in the recent IARC evaluation of carcinogenicity of ethylene oxide within the Monographs programme (IARC, 1994).

The utility of using adducts as markers of biologically effective dose is limited in case–control studies due to the relatively short half-life of most adducts evaluated to date. They may have utility in prospective cohort studies, again with the caveats previously described for all exposure markers.

Biomarkers of somatic cell mutations

Somatic mutations provide evidence of irreversible genetic damage (Albertini & Hayes, this volume). Furthermore, specific mutations (mutation spectra) may, in principle, identify exposures to specific agents or mechanisms; however, measurements of mutations are usually less sensitive in this regard than are other biomarkers of exposure (e.g. metabolites, adducts). Somatic mutations should have their greatest utility in epidemiological studies when it is possible to establish a qualitative association between exposure and specific mutations, thus identifying the agents of concern. They also have promise as surrogate markers of outcome. It must be emphasized, however, that for this last purpose, mutations must initially be assumed to be associated with increased risk of cancer. In-vivo mutational response may also be used

to define interindividual differences in sensitivity to mutagens or carcinogens.

Somatic mutations have potential application in case–control studies when the exposures of concern are chronic and stable, and the disease does not alter mutagenic responses in the test cells. Their greatest utility, however, may be in studies that use mutations as an intermediate end-point to evaluate genetic responses under various exposure conditions and in studies concerned with prevention. There is a need in the future for rapid, inexpensive and sensitive molecular assays that may increase the range of genes and tissues amenable for study; this, in turn, should facilitate the evaluation and potential use of those biomarkers as surrogates of disease-relevant genotoxic responses.

Biomarkers of cytogenetic damage

Chromosomal aberrations are commonly found in tumour cells and, in some cases, there is very good evidence that they play a causative role in carcinogenesis (Tucker *et al.*, this volume). Two prospective cohort studies have shown that chromosomal aberrations measured in peripheral lymphocytes were associated with an increased risk of cancer (Hagmar *et al.*, 1994; Bonassi *et al.*, 1995). Non-target cells may be appropriate tissue for this marker. Translocations, in particular, are useful for quantifying certain types of acute and chronic exposure, as well as exposure that occurred many years previously. Translocations are therefore a reasonable and appropriate biomarker for use in cross-sectional studies because they provide a direct and quantifiable indication of DNA damage. In case–control studies, the analysis of stable translocations in early, local disease may provide an insight into accumulated cytogenetic damage and its link with cancer. It is possible, however, that the frequency of translocations is affected by advanced disease, but this remains to be established. The use of translocations in prospective cohort studies could have substantial utility but is limited by the logistic constraints of culturing or cryopreserving lymphocytes from a large number of subjects.

The analysis of translocations by fluorescent in-situ hybridization (FISH) with probes that paint whole chromosomes represents an important advance in cytogenetics; FISH is a rapid, sensitive and highly reproducible technique which relies on the detection of colour 'junctions' between painted and unpainted chromosomes. Improvements in the techniques of molecular cytogenetics continue to be made, and new DNA probes are being developed on a regular basis. Cytogenetic assays have been performed on metaphase cells, but the more recent applications of FISH to interphase cells means that cell culture is no longer required and that relevant target tissues can be examined directly for chromosome rearrangements. The analysis of interphase cells appears to have significant promise for use in future epidemiological studies and may make the banking of the relevant tissues from a large numbers of subjects feasible.

Biomarkers of genetic susceptibility

Biomarkers have increasingly played a role in studies that evaluate the role of genes in cancer (Caporaso & Goldstein, this volume). Family-based studies have led to the mapping of several cancer-related genes characterized by high penetrance, high absolute and relative risk, but low attributable fraction for the most common cancers. The environment plays a variable but arguably small role in the manifestation of these genes. Other genes (e.g. metabolic polymorphism genes that lie in the metabolic pathways of carcinogens and therefore have a mechanistically plausible role) are characterized by modest relative risks, low absolute risks and high attributable risks. The environment is crucial in determining the effects of these genes. A major contrast between these two is the low prevalence of the former and the high prevalence of the latter.

For a given genotype, the phenotype is determined by the penetrance. The use of phenotype or genotype in a specific study depends on numerous factors. In general, the methodology of genotyping has increased because of rapidly improving technology. Recent advances in the analysis of genetic polymorphisms using DNA present in serum, in formalin-fixed, paraffin-embedded pathology samples and in material collected non-invasively by buccal swabs and oral rinses will provide new opportunities for incorporating biomarkers of genetic susceptibility into epidemiological studies.

Tumour biomarkers

Tumour markers include characteristics of tumours at the anatomical, histological, serum, chromosomal

and molecular levels (Zhang *et al.*, this volume). Using tumour markers in the assessment of etiological heterogeneity represents a progression from these previous studies and may have greater specificity for particular patterns of exposure. Tumour markers can readily be incorporated into case series and case–control studies, as well as into prospective studies that study the relationship between biomarkers in an initial tumour and the risk of developing second primary tumours or the relationship between precursor lesions and subsequent malignancies. There is a rapidly increasing ability to evaluate a range of tumour markers in formalin-fixed, paraffin-embedded tissue. Frozen tissues are easier to assay, but more difficult and costly to obtain and store for epidemiological purposes.

Evaluating sources of variation in biomarkers

Before embarking on an epidemiological study that uses a biomarker, it is important to understand the potential measurement error in the biomarker (Gompertz, this volume; Vineis, this volume; White, this volume). The researcher's first concern should be to prevent or rule out, as far as is practicable, differential measurement error. Because differential measurement error can bias the odds ratio in either direction, its presence to any appreciable degree in a biomarker will invalidate its use in an epidemiological study. Differential measurement error is a particular concern in case–control studies and among the early cases in cohort studies when the marker may be influenced by pre-clinical disease, by the effects of the disease after diagnosis or by treatment.

When differential measurement error is not likely to be present, the researchers should focus on assessment of the non-differential error, or at least some of the major components of error in the biomarker. Error components include laboratory variation, variation from specimen collection and storage, and biological variation. The latter may be short-term (e.g. day to day), medium-term (e.g. seasonal), or long-term (e.g. variation over an etiologically relevant period of years). Ideally, one would conduct a validity study in which the biomarker to be used was compared to a perfect (true) measure. However, much of the total error can be measured in a well-designed reliability study which requires the collection and analysis of two (or more) specimens from a group of subjects in a way that the error in one measure is not repeated in another, e.g. the two specimens are collected at different times over the relevant etiological time period, and handled, stored and analysed with the variation in specimen collectors, laboratory technicians or batches that would occur in the parent epidemiological study.

Similarly, reliability studies could be designed to partition the specific components of error: the effects of handling and storage; laboratory variation; and short-, medium- and long-term biological variation. In any case, researchers should not assume that small laboratory error implies that a measure is good, because these other sources of error introduced by the design and needs of the epidemiological study can be far greater than the laboratory measurement component of error.

Storage of biological samples

The use of biological markers in epidemiological studies often requires storage of the relevant biological samples for a period of time, which may vary from a few weeks up to years or even decades, depending on the study design (i.e. cross-sectional, prospective) and the timing of the laboratory analysis (Landi & Caporaso, 1997, this volume). Factors that can be considered in the choice of storage method include the type of biological material; the type of laboratory analysis planned; the duration of storage and spread over time of laboratory analysis; and the logistical and practical conditions. For long-term storage of a large collection of blood samples, the safest and, in most cases, the most efficient method of storage is at liquid nitrogen temperature.

Storage temperature should be as low as possible, depending on the biological material, and should ensure stability of all potential analyses over long periods of time. For very long-term storage, it is worthwhile to store pools of biological specimens containing known concentrations of the analytes of interest in order to be able to monitor possible degradation over time. A stable isotope standard may be useful in such cases.

Application of biomarkers

Biomarkers may be used in epidemiological studies to increase the information obtained from classical study designs and to expand the areas of scientific inquiry to which epidemiology can contribute. Use of

biomarkers in epidemiological research has grown significantly in recent years and can be expected to increase rapidly in the years ahead. The areas of application include research on etiology of disease; correlation between external exposure and internal dose; susceptibility and gene–environment interaction; clinical trials; disease mechanisms and pathogenesis; and cancer prevention.

Etiological epidemiological research

In etiological research, biomarkers of exposure represent the independent variable and are used to predict disease occurrence. Research of this type has been limited in the past because of the need for transitional studies to develop, characterize and evaluate biomarkers in their application. When biomarkers are tested as outcome variables they are useful in often being much more prevalent in the population than cancer itself and they occur much closer in time to the exposure than does cancer; exposure will therefore be measurable more accurately by whatever method is used. The majority of epidemiological biomarker research reported to date has been from transitional studies. As this body of literature has accumulated, it is now becoming possible to use selected biomarkers in the context of classical etiological research.

At current levels of development of biomarkers, there will still be many circumstances in which questionaire methods of assessing exposure are complementary or preferable to technically sophisticated biomarkers of the same exposure. Assays of urinary levels of metabolites of tobacco components such as nicotine, for example, reflect only short-term consumption. Such assays tell us about smoking habits in the relatively short time before measurement; questionnaires are still preferable to determine long-term exposure.

Etiological studies using biomarkers generally employ case–control or cohort epidemiological designs. Modifications of these are the case–case design and the applied transitional study. The case–control study nested within a cohort offers distinctive opportunities by using blood or tissue collected in the past.

The use of biomarkers in case–control studies may necessitate greater use clinic-based (or hospital-based) studies (Potter, this volume). Clinic-based studies of biomarkers may be more feasible, practical and economical. They also have a higher response rate with respect to collection of biological samples than community-based studies, and differential participation by cases and controls is less of a problem. For case–control studies of precursor lesions, no population-based registries exist from which to identify cases, whereas such cases can be identified in the health care setting and, given appropiate survey conditions, in random samples of the population. In cohort studies, where the biomarker is the exposure of interest, unexposed subjects, defined as people not carrying the biomarker, may be identified through the health care setting. Tissue specimens may be available only through pathology departments associated with hospitals. Opportunities to obtain specimens may be enhanced if health care providers are co-investigators in the research. The method for selecting study subjects is an important potential source of bias in clinic-based studies. In studies based on biomarkers, however, it may be less important than in other types of epidemiological studies, because the characteristics of the subject measured by the biomarker (e.g. genetic polymorphism status) may not be linked to the probability of inclusion in the study. Similarly, it is unlikely that markers of exposure and effect are affected by referral and detection, since the individual's biomarker status prior to the study is not known. In most cases, however, it is unknown how subjects are selected with respect to the biomarker, and the possibility of selection bias should always be considered.

Exposure–internal dose and exposure–biological response correlations

Transitional studies of exposure frequently correlate external exposure measured in a traditional fashion (e.g. interview or ambient monitoring) with internal dose measured in body fluids or tissues. These studies are generally cross-sectional or short-duration longitudinal in design. The exogenous measure is considered the independent variable, and the internal dose the dependent variable. Rather than an internal dose measure, the dependent variable can be a biological response marker, e.g. a *p53* signature mutation associated with aflatoxin exposure (Links *et al.*, 1995). Such transitional studies frequently focus on heavily exposed (and unexposed) persons in order to minimize sample size and maximize the opportunity

to identify and characterize the biomarker. If the correlation between the two types of exposure measurements is high and can be quantified with good precision, future studies may be able to employ the more easily obtained and inexpensive exposure measurement.

Susceptibility and gene–environment interaction

Studies of susceptibility and gene–environment interaction are emerging as an important component of epidemiological research (Garte *et al.*, this volume). This represents a significant development in epidemiology; it emphasizes the importance of individual differences and the combined influence of genes and environment in determining disease risk. Although relevant genes lie in a continuum with respect to prevalence and expressivity, a distinction should be made between single gene mutations of large effect and polymorphisms of small effect. *BRCA1* exemplifies the former, and metabolizing enzymes for substrates that form carcinogenic metabolites illustrate the latter. Intermediate situations exist in individuals heterozygous for some genes, such as ataxia telangiectasia, which may affect a large proportion of the population, and for which there is some evidence of increased risk of various cancers, including breast cancer (Easton, 1994). Because estimates of cancer risk in relation to single gene mutations have been derived from high-risk families, it is not yet known what the cancer risk is for gene carriers within the general population. There is a need for epidemiological studies of cancer in the general population to study risk in single gene carriers who are not members of high-risk families.

Several methodological issues have arisen in the context of these studies. Lack of concordance between genotype and phenotype has been reported and it is not easy to discern which is more informative with respect to disease effects. However, some advantages of studies of susceptibility and gene–environment interactions are the potential for case–case studies (Begg & Zhang, 1994) and for studying 'exposed' subjects only.

Khoury *et al.* (1988) has described six models for gene–environment interaction. These concepts may be simplified by considering the possibility of a main effect and effect modification for both the susceptibility marker and the exposure. For some genes, the susceptibility marker can behave as an 'exposure', increasing risk of disease in the entire study population. Usually, however, susceptibility is manifested as an effect modifier of exposure. If the population is stratified into two groups, based on presence or absence of the susceptibility gene, the effect of the exposure variable may be more evident in the susceptible group. It may also be useful to stratify on levels of the exposure variable, e.g. smoking or age groups, to identify qualitative and quantitative aspects of the association between the susceptibility marker and disease across levels of the exposure.

Cancer prevention

In cancer epidemiology, biomarkers can be used to offer quantitative insights about biological events occurring at different stages of the pathologic process (McMichael & Hall, this volume). The improved knowledge of the natural history of cancer should have application to research aiming at cancer prevention and control. At least three aspects of cancer prevention or cancer prevention research may make use of biological markers. They are screening, community-based intervention trials and the monitoring of biomarkers as risk factors for disease. Screening may concentrate on the identification of precursor lesions or early stage disease (e.g. cervical cytology), high-risk individuals (e.g. prostate-specific antigen, PSA) or susceptibility markers (e.g. *BRCA1*). Cervical cytology is well established as an effective tool to reduce mortality from invasive cervical cancer. PSA is controversial because it is not specific to invasive prostatic cancer or its precursors. It cannot distinguish between precursor lesions that will invade and the majority that will not, resulting in a significant amount of over-treatment and morbidity. Furthermore, and crucially, there is no evidence that screening reduces mortality from prostatic cancer. Screening for *BRCA1* became possible only recently when the gene was cloned and sequenced (Futreal *et al.*, 1994; Miki *et al.*, 1994). Currently, testing for *BRCA1* has been suggested only for women in high-risk families in whom the gene is known to segregate, or for women with multiple first-degree relatives who have had an early age at diagnosis of breast cancer. Even in these situations, screening is not universally accepted because of the limited risk and disease management options, which markedly reduce the utility of the genetic information. Even

in the cases of high penetrance cancer genes, however, there is the possibility to contribute to cancer prevention by modifying the environmental factors that interact in the carcinogenic process.

Community-based intervention trials for risk reduction may benefit from biological markers. For example, smoking cessation trials have found that urinary cotinine levels may be a better indicator of smoking status, when measuring the trial outcome, than reports on interview or questionnaire.

Monitoring biomarkers of exposure to carcinogenic agents in populations may reveal trends and assist in risk assessment in high-risk groups.

Clinical prevention trials

The use of biomarkers in chemopreventive trials raises important methodological issues. The testing of chemopreventive agents in clinical trials using the end-point of cancer incidence requires a study period of many years, very large sample sizes and great expense. Therefore, short-term, smaller clinical trials that use surrogate end-point biomarkers (SEB) have had to be developed. A SEB may be defined as an early change during the intraepithelial, pre-invasive phase of neoplastic progression, at the molecular, cellular or tissue level, whose response to a chemopreventive agent predicts the effectiveness that the agent would have in a large clinical trial using the end-point of cancer incidence reduction (Boone & Kelloff, this volume). Examples of SEBs include computer-assisted quantitative image analysis of pathological specimens, which measures nuclear features (altered size, shape, and chromatin texture), and cytological features by binding to chromagen–antibody conjugates. Markers and indices of proliferation (e.g. Ki-67, PCNA), oncogene mutation or amplification, and allelic loss and other alterations may also be useful SEBs.

Studies are needed, however, to evaluate the validity of SEBs. To be suitable for use as an SEB, there should be evidence that the marker is a necessary step on the pathway to cancer, or at least very highly correlated with cancer occurrence, i.e. the attributable fraction for the marker in relation to cancer must approach 1. This evidence may be obtained primarily from observational epidemiological studies and clinical trials that incorporate the marker and have explicit cancer end-points. Certain laboratory and pathology studies may also yield pertinent evidence, as in the case of the colon adenoma to carcinoma sequence. With markers having attributable fractions substantially less than 1, there may exist alternative pathways to cancer that bypass the marker in question. An exposure or intervention may operate through the alternative pathway(s) in a way that offsets the cancer effect mediated by the original marker. For such markers, inferences to cancer are problematic (Schatzkin *et al.*, 1990, and this volume).

Although there are well-recognized difficulties in extrapolation of carcinogenesis and chemoprevention data from animal experiments to humans, animal models provide opportunities for biomarker research. Linkages between mechanistic pathways in different species may ultimately be useful in resolving problems in inter-species extrapolation.

Ethical aspects of biomarkers in cancer epidemiology

The use of biological markers presents potential ethical issues because biomarkers are obtained from an individual's unique tissues and can be used to provide important data about exposures, biological effects and susceptibility to cancer (Schulte *et al.*, this volume). The ethical issues arise from the possibility of abuse or misuse of biomarker data and failure to respect the rights of persons participating in research. Biomarker data can be misused by failing to keep data confidential and by using it, or allowing it to be used, to stigmatize research subjects. Research subjects have a right to privacy as well as a right to be told of risks of participating in biomarker studies and of any clinically important findings. There remains a range of opinion whether banked specimens collected for one study or purpose may be used for another. Generally, the degree of risk in participation and the extent to which the results are clinically important should be considered when using banked specimens and determining whether subjects need to be notified.

Recommendations for future studies of cancer epidemiology involving biomarkers

The Workshop considered the main issues regarding the use of biomarkers in future studies in cancer epidemiology, and agreed a set of recommendations. These are presented in Box 1.

Box 1. Recommendations for future studies

1. Biomarkers should be chosen because of their biological relevance to the question and on the basis of the science, not the technology.

2. Biomarkers should be evaluated by transitional studies before use in full-scale studies. Regulatory and funding agencies should become aware of the importance of transitional studies for the conduct of epidemiological studies based on biomarkers.

3. In addition to etiological research, appropriately validated biomarkers should be used in prevention: in screening, as biomarkers of neoplasia or susceptibility, in clinical trials (e.g. biomarkers of dietary modification and compliance, and biomarkers of surrogate end-points), and in monitoring exposure to carcinogenic agents (e.g. trends in biomarker status in general populations and subgroups, exposure status and risk assessment in high-risk populations, and exposure/susceptibility interactions).

4. Epidemiological studies based on biomarkers should satisfy the requirements of good epidemiological research design, including identification and minimization of potential sources of bias and confounding. Research should be conducted on methodological aspects to clarify advantages and disadvantages of different options in the design of biomarker-based studies, in particular to determine wehther, in the clinic-based (or hospital-based) setting, studies of biomarkers of exposure or susceptibility are less subject to selection biases than traditional studies of behavioural and environmental risk factors of disease.

5. Studies should be designed with appropriate statistical power to minimize the problem of random error and maximize the capacity to establish the presence and strength of interactions, particularly between markers of susceptibility and exposure.

6. It is important to define the components of variance in order to establish the capacity of the biomarker to measure, as precisely as possible, the relevant exposure, process and outcome. For continuous measurements, it is highly desirable that the amount of random error due to within-person variability in the measurement be estimated in a subset of the study population. If this cannot be estimated directly in the study population, a similar population should be used because the ratio of within-person to between-person variance may vary substantially between populations due to differences in the range of exposure, or fluctuations in exposure.

7. There should be strong and open communication between laboratory scientists and epidemiologists with regard to identification, investigation and control of specific sources of measurement error in the biomarker assay. Such issues would include the need for replicate measurements and standardized non-variable experimental protocols during the study; internal controls; operator blindness; monitoring experimental drift; sample handling; and development of study design by both epidemiologists and laboratory scientists to deal with these issues.

8. Epidemiologists should work closely with laboratory investigators when developing protocols for the collection, processing and storage of biological samples to ensure that the samples can be analysed for the main biomarkers of interest and that the requisite quality control procedures are implemented.

9. Standardized procedures should be developed for the collection, handling, processing and storage of biological samples. A biospecimen documentation sheet should be generated for each biological sample collected, documenting critical variables that may affect biomarkers.

10. Storage of blood samples in multiple aliquots, and ideally in separate freezers, is advisable whenever long-term storage and multiple measurements spread over long time periods are planned. The samples should be stored at the lowest feasible temperature. Assessment of sample stability would be facilitated by storage of aliquots of a pool prepared specially for this purpose.

11. Samples should be collected not only for the main biomarkers of interest in a particular study; rather, they should be processed and stored in a way that allows a wide range of biomarkers to be tested in the future. In particular, consideration should be given to collecting a source of genomic DNA in epidemiological studies to allow the evaluation of genetic markers. While this is optimally obtained from peripheral blood samples, various non-invasive methods for

Box 1. (Contd) Recommendations for future studies

collecting genomic DNA are available and continue to be developed, including the use of buccal swabs or oral rinses to collect exfoliated epithelial cells.

12. Where feasible, tumour pathology samples should be collected for analysis of tumour characteristics in addition to histological confirmation. This generally requires collecting more material than is usually obtained for histological evaluation alone. If tumour blocks are available, collection of multiple slides from formalin-fixed, paraffin-embedded blocks is desirable. Availability of frozen fresh tumour and normal tissue should be encouraged when relevant and feasible.

13. In studies of the natural history of malignant disease and its prevention, it is recommended that molecular and protein alterations be studied in normal tissues, benign lesions, premalignant precursor lesions and the surrounding unaffected tissue. This evaluation should also be extended to non-target tissues and to the other organ, in the case of paired organs. Assays that can be applied to paraffin-embedded, formalin-fixed tissues are of particular importance to epidemiological research.

14. Special emphasis should be given to improving correlation between animal studies and human in-vitro studies designed to further develop, refine and allow for better biological interpretations of results obtained from epidemiological studies using biomarkers. An iterative process is envisaged in which epidemiological and laboratory components collaborate in efforts to define pathogenesis.

15. Care should be taken to consider the possible ethical issues when using biological markers. Biological marker data should be kept confidential and not released in ways that will allow for identification of participants in research.

16. When using banked specimens, the need for informed consent and reporting back results to subjects should be based on the degree of clinical risk associated with the biomarker.

17. Investigators should be encouraged to publish, or make publicly available, data involving biomarkers even if no associations are observed. This applies in particular to transitional studies.

18. More opportunities for joint training in both laboratory and epidemiological methods should be developed. Encouraging effective collaboration between laboratory-based scientists and epidemiologists and the creation of interdisciplinary programmes would foster research in this area.

References

Begg, C.B. & Zhang, Z.F. (1994) Statistical analysis of molecular epidemiology studies employing case-series. *Cancer Epidemiol. Biomarkers Prev.*, 3, 173–175

Bonassi, S., Bolognesi, C., Abbondadolo, A., Barale, R., Bigatti, P., Camurri, L., Dalpra, L., DeFerrari, M., Forni, A., Lando, C., Padovani, P., Pasquini, R., Stella, M. & Puntoni, R. (1995) Influence of sex on cytogenic end points: evidence from a large human sample and review of the literature. *Cancer Epidemiol. Biomarkers Prev.*, 4, 671–679

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V. & the IBSCC Study Group (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J. Natl Cancer Inst.*, 87, 796–802

Cole, P. & MacMahon, B. (1969) Oestrogen fractions during early reproductive life in the aetiology of breast cancer. *Lancet*, i, 604–606

Dickinson, L.E., MacMahon, B., Cole, P. & Brown, J.B.(1974) Estrogen profiles of Oriental and Caucasian women in Hawaii. *N. Engl. J. Med.*, 291, 1211–1213

Easton, D.F. (1994) Cancer risks in A-T heterozygotes. *Int. J. Rad. Biol.*, 66, S177–182

Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J.D., Marks, J., Ballinger, D.G., Barrett, J.C. Skolnick, M.H., Kamb, A. & Wiseman, R. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science*, 266, 120–122

Green, R. & Jacobsen, D.W. (1995) Clinical implications of hyperhomocysteinemia. In: Bailey, L.B., ed., *Folate in Health and Disease*. New York, Marcel Dekker Inc. pp. 75–122

Hagmar, L., Brøgger, A., Hantseen, I.L., Heim, S., Högstedt, B., Knudsen, L., Lambert, B., Linnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerfving, S. & Sorsa, M. (1994) Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic Study Group on the Health Risk of Chromosome Damage. *Cancer Res.*, 54, 2919–2922

Hulka, B.S., Wilcosky, T.C. & Griffith, J.D., eds (1990) *Biological Markers in Epidemiology*, New York, Oxford University Press

Hulka, B.S., Liu, E.T. & Lininger, R.A. (1994) Steroid hormones and risk of breast cancer. *Cancer*, 74 (Suppl. 3), 1111–1124

Institute for Environment and Health (1996) IEH Report on the Use of Biomarkers in Environmental Exposure Assessment, (Report R5), Leicester, Institute for Environment and Health

IARC (1994) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, Vol. 60, *Some Industrial Chemicals*, Lyon, pp. 73–159

International Collaborative Group (1982) Circulating cholesterol levels and risk of death from cancer in men aged 40 to 69 years. *J. Am. Med. Assoc.*, 248, 2853–2859

Khoury, M.J., Adams, M.J., Jr & Flanders, W.D. (1988) An epidemiological approach to ecogenetics. *Am. J. Hum. Genet.*, 42, 89–95

Kuopio, T., Ekfors, T.O., Nikkanen, V. & Nevlainen, T.J. (1995) Acinar cell carcinoma of the pancreas: report of three cases. *APMIS*, 103, 69–78

Lemaster, G.K. & Schulte, P.A. (1993) Biologic markers in the epidemiology of reproduction. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practice*, San Diego, CA, Academic Press. pp. 385–406

Links, J.M., Kensler, T.W. & Groopman, J.D. (1995) Biomarkers and mechanistic approaches to environmental epidemiology. *Ann. Rev. Pub. Health*, 16, 83–103

MacMahon, B., Cole, P., Brown, J.B., Aoki, K., Lin, T.M., Morgan, R.W. & Woo, N.C. (1974) Urine oestrogen profiles of Asian and North American women. *Int. J. Cancer*, 14, 161–167

MacMahon, B., Trichopoulos, D., Brown, J., Andersen, A.P., Cole, P. deWaard, F., Kauraniemi, T., Polychronopoulou, A., Ravnihar, B., Stormby, N. & Westlund, K. (1982) Age at menarche, urine estrogens and breast cancer risk. *Int. J. Cancer*, 30, 427–431

MacMahon, B., Cole, P., Brown, J.B., Paffenbarger, R., Trichopoulos, D. & Yen, S. (1983) Urine estrogens, frequency of ovulation and breast cancer risk: case–control study in premenopausal women. *J. Natl Cancer Inst.*, 70, 247–250

McMichael, A.J., Jensen, O.M., Parkin, D.M. & Zaridze, D.G. (1984) Dietary and endogenous cholesterol and human cancer. *Epi. Reviews*, 6, 192–216

Mendelsohn, M.L., Peeters, J.P. & Normandy, M.J., eds. (1995) *Biomarkers and Occupational Health: Progress and Perspectives*, Washington, DC, Joseph Henry Press

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R. Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K.,

Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C ., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A. & Skolnick, M.H. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266, 66–71

Muñoz, N., Bosch, F.X., de Sanjosé, S., Tafur, L., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C., Ascunce, N., Gonzalez, L.C., Kaldor, J.M., Guerrero, E., Lörincz, A., Santamaria, M., Alonso de Ruiz, P., Aristizabal, N. & Shah, K. (1992) The causal link between human papillomavirus and invasive cervical cancer: a population-based case–control study in Colombia and Spain. *Int. J. Cancer*, 52, 743–749

Nakazawa, H., English, D., Randell, P.L., Nakazawa, K., Martel, N., Armstrong, B.K. & Yamasaki, H. (1994) UV and skin cancer: specific p53 gene mutation in normal skin as a biologically relevant exposure measurement. *Proc. Natl Acad. Sci.*, 91, 390–364

Qian, G.S., Ross, R.K., Yu, M.C., Yuan, J.M., Gao, Y.T., Henderson, B.E., Wogan, G.N. & Groopman, J.D. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol. Biomarkers Prev.*, 3, 3–10

Ross, R.K., Yuan, J.M., Yu, M.C., Wogan, G.N., Qian, G.S., Tu, J.T., Groopman, J.D., Gao, Y.T. & Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, 339, 943–946

Schatzkin, A., Freedman, L.S., Schiffman, M.H., & Dawsey, S.M. (1990) Validation of intermediate end points in cancer research. *J. Natl Cancer Inst.*, 82, 1746–1752

Schulte, P.A., & Perera, F.P., eds (1993) *Molecular Epidemiology: Principles & Practices*, San Diego, Academic Press

Sherwin, R.W., Wentworth, D.N., Cutler, J.A., Hulley, S.B., Kuller, L. & Stamler, J. (1987) Serum cholesterol levels and cancer mortality among 361,662 men screened for the multiple intervention trial. *J. Am. Med. Assoc.*, 257, 943–947

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Transitional studies

P.A. Schulte and F.P. Perera

Transitional studies are studies using biological markers that bridge the gap between laboratory experiments and population-based epidemiology. The goal of these studies is to characterize and validate biomarkers and to assess the following: intra- and inter-subject variability; the feasibility of marker use in field conditions; confounding and effect-modifying factors for the marker; and mechanisms reflected by the biomarker. Another goal is to optimize the conditions for the use of biomarkers. Transitional studies involving biomarkers of exposure or effect are distinguished from etiological studies because the biomarker is generally the outcome or dependent variable. Despite this difference, transitional studies can be epidemiological studies, but they may also include laboratory studies to assess reliability (and accuracy) and to identify parameters for collecting, processing and storing biological specimens prior to assay. Generally, transitional studies involve healthy people, patients or workers with specific exposures. At some point in the validation of a biomarker the line between transitional and etiological studies becomes blurred. None the less, it is useful to identify transitional studies as a distinct set of efforts to validate and characterize biomarkers. Transitional studies can be divided into three functional categories: developmental, characterization and applied studies.

Transitional studies have been described as studies using biological markers that bridge the gap between laboratory experiments and population-based epidemiology (Hulka, 1991; Schulte, 1992; Schulte *et al.*, 1993; Rothman *et al.*, 1995). The goal of these studies is to characterize and validate biomarkers and to assess the following: intra- and inter-subject variability; the feasibility of marker use in field conditions; confounding and effect-modifying factors for the marker; and mechanisms reflected by the biomarker. Another goal is to optimize the conditions for use of biomarkers. Transitional studies involving biomarkers of exposure or effect are distinguished from etiological studies because the biomarker is generally the outcome or dependent variable. Despite this difference, transitional studies can be epidemiological studies, but they may also include laboratory studies to assess reliability (and accuracy, if possible) and to identify parameters for collecting, processing and storing biological specimens prior to assay. Generally, transitional studies involve healthy people, patients or workers with specific exposures. At some point in the validation of a biomarker the line between transitional and etiological studies becomes blurred. None the less, it is useful to identify transitional studies as a distinct set of efforts to validate and characterize biomarkers.

Transitional studies can be divided into three functional categories (Table 1): developmental, characterization and applied studies (Schulte *et al.*, 1993; Rothman *et al.*, 1995). There is no clear demarcation between the types of transitional study, but the categories are useful in identifying the clustering of preparatory research that is needed before a biomarker is ready for population research. All three types of transitional study are efforts to determine aspects of the validity of a biomarker.

Transitional studies use epidemiological methods in the development, testing and validation of biomarkers. These studies represent preparatory efforts to determine the parameters, limitations or characteristics of a biomarker prior to its use in etiological, prevention or other kinds of intervention efforts. Transitional studies build on scientific knowledge from various types of laboratory studies, including tissue culture and animal studies. The conduct of transitional studies is not a one-time step in the development of biomarkers but can represent steps in an iterative process. Thus a biomarker may be developed in the laboratory, validated in a transitional study using epidemiological methods, and then applied in an etiological study that raises new questions that stimulate new laboratory and transitional studies.

Table 1. Types of transitional studies

Type	Description
Developmental	Builds on laboratory and *in vitro* animal studies and involves the test of assay in humans
Characterization	Examines the variability of biomarkers in population subgroups and factors that could confound associations between markers and what they represent
Applied	Assesses the relationship between a marker and an underlying event such as an exposure, a biological effect or susceptibility

Overview of the validation process

Validation of biomarkers involves the clarification of factors that influence the ability of the marker to predict exposure or outcome in a test population. The issues covered here are discussed by Schulte & Perera (1993) and Perera & Mooney (1993). During the validation process, evidence is weighed to determine whether a biomarker measures what it claims to measure, be it exposure, disease or susceptibility. A more epidemiological definition is that a biomarker is valid to the extent that it measures the true marker populations, that is, with no measurement error (see White, this volume).

As shown in Table 2, validation can be considered to be a two-stage process, although many of the steps require iteration as new information becomes available. The ultimate goal of the validation effort is to allow biomarkers to be used in appropriate epidemiological and clinical applications, which will require a selection of the best biomarkers based on the criteria described below.

The first step is laboratory validation, in which the shape of the dose–response curve, low-dose sensitivity (the ability to detect the exposure at levels low enough to be of biological interest), exposure specificity, and reproducibility of the assay are tested. In the second step, known as epidemiological validation, the population sensitivity and specificity, intra-individual and interindividual variation in response and persistence, predictive value, and biological relevance and feasibility are evaluated.

In this validation process, pilot studies in high-dose groups are useful. Molecular epidemiological studies to monitor the carcinogenic potential of

Table 2. Criteria for molecular epidemiological validation of biomarkers

Laboratory	Epidemiological
Dose–response curve	Population sensitivity and specificity
Detection limit or low-dose sensitivity	Intra-individual variation over time:
Exposure specificity	a. without altering exposure
Reliability of the assay:	b. when exposure is removed (persistence/half-life) or changed
a. from run to run	Inter-individual variation:
b. from day to day	a. response to a given exposure
c. from one laboratory to another	b. persistence of biomarker
Optimal conditions for sample collection, processing and storage	Half-life in surrogate tissues
	Positive predictive value (yield of high-risk people)
	Feasibility:
	a. amount and availability of tissue
	b. cost
	c. time required for each assay
	Biological relevance to disease

Source: Perera & Mooney, 1993

industrial chemicals in workers have been a useful approach to validation because occupational exposures tend to be higher and more controlled than ambient exposures (Perera & Weinstein, 1982; Perera & Santella, 1993). Initial research linked a number of occupational exposures to increases in adducts, mutation and other biomarkers in workers. State-of-the-art studies have broadened to include assessment of smoking and dietary factors (e.g. vitamins, fat, oxidative damage), inherent factors such as genetic predisposition to biotransformation and repair capacity, and endogenous levels of enzymes and promoters.

Interdisciplinary collaboration

Effective collaboration between laboratory scientists and epidemiologists is critical in conducting transitional studies. These studies involve the first field testing of the markers and they require both laboratory and epidemiological expertise. The interaction of different disciplines requires that attention be paid to the underlying assumptions, paradigms and language of the various disciplines. Epidemiologists generally speak in terms of groups and risk to groups. Laboratory scientists tend to focus on individuals or components of an individual. Epidemiology is an observational science whereas laboratory disciplines use controlled experimental designs. Epidemiologists and laboratory researchers use the same word to mean different things. For example, when laboratory researchers speak of a valid marker, they are referring to the characteristics of the assay for the marker, whereas when epidemiologists speak of a valid marker, they are generally referring to one with a high predictive value or correlation with exposure or disease.

The scale of transitional studies may be a hindrance to effective collaboration. Laboratory research generally takes place on a small scale, whereas epidemiological studies can produce vast quantities of specimens. Even in a study that is small by epidemiological standards, e.g. of 30 people, the volume of specimens is often more than the research laboratory is used to processing. The laboratory workers can easily feel that they are not engaged in true scientific research but are merely providing a service. Meanwhile, the epidemiologist wonders why the laboratory is not able to handle the volume of specimens required for the study. (Wilcox, 1995).

Another hindrance to good interdisciplinary collaboration is the tension between assay consistency and assay improvement (Rothman, 1993; Wilcox, 1995). The relatively long term for epidemiological studies may span a period when improvements are made in the assay. The laboratory collaborators will naturally want to take advantage of the improved assays and incorporate them in the study without consulting the epidemiologist. This can be disastrous from the epidemiological point of view. The use of two different assay methods can ruin an epidemiological study. As new and more reliable assays are developed, scientists will be persuaded to replace previous essays. When a marker assay is new, measurements may differ from those in previous assays for the same marker. Although it may be reasonable to wait until laboratory techniques and estimates of variability display consistency, it is not reasonable or feasible to wait until a technique is so standard that no refinement is likely to occur. However, before modifying a technique during the course of a study, the old versus new results must be carefully evaluated for overall comparability.

Developmental studies

Reliability

When a candidate biomarker is identified in the laboratory, some very basic issues need to be solved before it can be considered for use in population studies. The first priority in evaluating a marker for use in population studies is to determine its reliability or reproducibility. As long as the assay is reliable, the ordering of subjects by the measure is preserved (Rothman *et al.*, 1995). Hence, there will be consistency within a study. Since this is all that is required for studying a marker–disease relationship, reliability and not accuracy is of primary importance (Rothman *et al.*, 1995). Reliability can be assessed optimally through analyses of blind replicate human samples that are representative of a range of values likely to be found in human populations.

Reliability encompasses both unsystematic random laboratory variation observed in repeated measurements and bias caused by non-random variation (Vineis *et al.*, 1993). To assess random error, multiple measurements are needed. Obviously, the random error in the arithmetic mean of several measurements is smaller than the random error of

an individual measurement. Quantitative indices of the extent of random variation of a biological marker can be used to determine whether the reliability of a given measure is sufficient for the purpose being considered. The two most common indices are the standard error of the measurement and the reliability coefficient (Massey, 1986).

Analysis of random laboratory variation involves several steps. First, multiple analytical measurements of the same biological specimen (or several biological specimens from the same individual at the same point in time) must be made to estimate the variability due to random analytical errors. This fraction of the total random variability of a biological marker is usually minor. Second, multiple measurements of a marker must be made for one individual over time to estimate the intra-individual temporal variability. Third, multiple measurements across different individuals must be made to estimate interindividual variability in the value of the marker. In the second and third cases, random error is a component of the variability but systematic errors may also contribute (e.g. circadian cycles in the value of a marker or differences among individuals due to genotype). Most molecular epidemiological research using biological markers seldom requires large numbers of individual measurements. Thus, a small number of individuals can be used as a sample of the infinitely larger population to which the distribution refers. The standard error indicates how the mean of that sample is distributed around the mean of the larger population. Hence, the standard error of the mean reflects the reliability of the sample mean as an indicator of the population mean (Massey, 1986). This value may not be as informative as the reliability coefficient for evaluating markers to be used in epidemiological studies.

The reliability coefficient is technically known as the intraclass coefficient of reliability (Shrout & Fleiss, 1979; Fleiss, 1986) and ranges from 0 to 1. If each measurement is identical, the intraclass coefficient is 1.0. The greater the variation among measurements, the lower is the reliability. Fleiss (1986) has evaluated the impact of unsystematic bias variation in measurement, described the problematic consequences of unreliability, and recommended how unreliability can be controlled. The consequences described by Fleiss (1986) include the need to increase sample size to reduce unreliability; the high rates of misclassification in studies of the association between exposure and disease; and the consequent underestimation of the association between a health measure and the measured extent of exposure to an environmental risk factor. All these factors pertain to studies using biological markers of exposure or effect. Fleiss (1986) recommends that unreliability be controlled by conducting pilot studies and replicating measurement procedures on each study subject.

Another aspect of developmental transitional studies is to define the optimal conditions for collecting, processing and storing biological specimens. The issues mentioned here are discussed by Winn & Gunter (1993) and in other chapters in this volume. The major lessons are that, at all stages of specimen handling, variation and/or error can be introduced, and careful attention is required to prevent these untoward aspects. In addition to the reliability of biomarkers, it is important to understand the biokinetics and stability aspects before application in population studies. These issues have been addressed in previous publications (Droz, 1993; Bernard, 1995). As discussed, valid biomarkers, particularly of exposure, will be those that have biological relevance, defined pharmacokinetics and temporal relevance. For markers of exposure, the parameters that best summarize the behaviour of a chemical in biological systems is the elimination half-life, which reflects both the affinity of the chemical for the biological matrix and the efficiency of excretory or metabolic processes of elimination. Bernard (1995) has suggested four categories for biomarkers of exposure: half-life less than 12 hours; half-life between 12 and 100 hours; half-life between 100 hours and 6 months; and half-life greater than 6 months. Generally, for etiological epidemiological studies, biomarkers in the latter two categories will be most useful.

Biological relevance

The biological relevance of a biomarker is of prime importance in the selection and validation of a marker. The validation of a specific biomarker must include consideration of its biological relevance to the disease or exposure under study and the position in the continuum between exposure and disease. Given the fact that many exposures cause multiple diseases, the validation of a biomarker's

relationship to a disease first requires a clear definition of the disease end-point. A well developed hypothesis of the relationship between the biomarker and other events in the exposure–disease continuum is necessary.

While knowledge of the stability and natural history are important before population application of biomarkers of effect, it is important to assess the attributable risk or proportion (Schatzkin *et al.*, 1990; Benichou, 1991; Trock, 1995). The attributable proportion associated with a particular biomarker is an estimate of the proportion of cancer cases that must progress through the biomarker. This is not simply the proportion of all cases that are positive for the biomarker, because the biomarker will occur in some cases ('background cases') even when the exposure of interest or the biological events associated with the biomarker are not etiological events for those cases (Trock, 1995). This assessment will help to identify the possible mechanism by which the biomarker is related to the causal pathway for the cancer and to determine the extent to which the biomarker truly represents a biological event intermediate between exposure and cancer.

Shatzkin *et al.* (1990) have developed a framework for considering exposure–marker–disease relationships:

- A single marker is known to be linked causally and, hence, is necessary and sufficient for disease.
- Multiple pathways and multiple biomarkers are identified and any marker is sufficient but not necessary for disease.
- An exposure leading to disease operates through an unobservable event which in turn lends to an intermediate marker that is not directly biologically related to the disease but which may correlate with its occurrence.

It is the role of developmental and applied transitional studies to refine the view of these various mechanisms and to attempt to confirm which mechanism describes the type of marker being considered. It is important to understand underlying mechanisms in order to distinguish measurement error problems from multiple pathways.

Determining the number of subjects

The number of subjects needed in transitional studies is a matter of judgment. If the objective is to determine simply whether an assay 'works', it is appropriate to test the assay on a small number of available subjects without regard for sample size or representativeness. However, in transitional studies to determine the validity of a marker and its variability, it is important that subjects are selected through defensible sampling designs. Critical in this regard will be the need to minimize selection bias and to generate adequate statistical power when null hypotheses of no difference among groups or perhaps no association among biomarkers are to be tested.

For biomarkers of exposure, it has been shown to be useful to select subjects from groups known to have high exposures (such as chemotherapy patients or workers in specific occupational exposure groups) to maximize exposure differences between groups and to obtain an indication of whether an assay will identify a potentially large 'signal', and ultimately a dose–response relationship. For biomarkers of effect, studies involving histologically defined subsets of patients are useful in determining the association of biomarkers and a particular cancer. The calculation of sample size for transitional studies is the same as for etiological studies. For a situation in which two groups (e.g. cases and controls) are compared on the basis of presence or absence of some biomarkers, the formulation for the calculation of sample size for a given combination of significance level, power and size of expected difference as measured in terms of relative risk is well known (Schesslman, 1982; Hertzberg & Russek-Cohen, 1993).

When groups of subjects are being compared for shifts in central tendency (such as mean values), the sample size is calculated according to a well known formula described in elementary textbooks that gives the sample size as a function of alpha, power, the size difference to be detected and the variance of marker levels within groups. When the investigator is interested in inferring the relative risk of a disease as a function of incremental change in the level of a given biomarker, the sample size is determined by a formulation involving type I and II error rates, mean levels in cases and controls, and variance in cases and controls, respectively (Hertzberg & Russeck-Cohen, 1993).

The determination of sample size reflects a tension between feasibility and cost issues on the one

hand and confounding on the other. What are the best design and analysis strategies to address these problems? For biomarkers of exposure it is essential to have an unexposed comparison group or subgroups with a range of exposures. Matching exposed and unexposed subjects on potential confounding factors can increase the precision of the measure of association between exposures and outcome, but this may decrease the study's power while controlling for confounding on the matching variables (Hulka, 1990). Alternatively, a study design could involve restriction of suspected confounders, i.e. limiting the subjects eligible to those with or without a particular potentially confounding factor. This may minimize confounding but will not allow for analysis of effect modification by the restriction variables.

In practice, the number of subjects will depend on both the statistical power and the cost of obtaining subjects, specimens and assays. Unlike more traditional epidemiological studies, in which unexposed persons may be more accessible and cheaper to recruit than exposed persons, the opposite situation may occur in transitional studies where there is little incentive for non-exposed or non-diagnosed persons to provide biological samples (Hulka, 1990).

Characterization studies

Once a marker has been sufficiently developed in terms of the reliability of the assay and optimal conditions for handling, it is necessary to assess its characteristics in human populations (Schulte *et al.*, 1993; Rothman *et al.*, 1995). The objective of these studies is to identify factors that are confounders or effect modifiers which should be taken into account in etiological or public health studies.

One type of characterization study involves the assessment of the frequency of a marker in various population subgroups, determined by characteristics such as age, race, sex, medical condition, behaviour, etc. These types of study should be conducted on a marker-by-marker basis depending on the underlying biology, mechanism and relation to various host factors.

The ultimate goal of characterization-type transitional studies is to assess interindividual variation and the genetic and acquired factors that influence the variation. Large interindividual variation will make it more difficult to predict the risk of disease if the responses of exposed and unexposed people overlap. Interindividual variation in response is thought to result from differences in a variety of factors, including personal exposure, ability to metabolize carcinogens, and differences in DNA repair, immune surveillance and nutritional status. Therefore, people with the same apparent exposure may vary widely in their response to carcinogens. For example, interindividual variation in the level of xenobiotics adducted to DNA results from processes that vary between people, affecting the concentration at the target tissue after distribution *in vivo*, metabolic activation and detoxification, capacity of repair, and persistence of the xenobiotic and cellular target *in vivo*.

A goal of molecular epidemiology is to elucidate the mechanisms that explain why people vary in their risk of disease. To do this, we must understand which part of the biological or biomarker measurement is due to laboratory variability or to intra-individual variation and which part reflects true interindividual variation. The mechanisms underlying interindividual differences in biomarker response are important because they are the same mechanisms that the body uses to mitigate a disease outcome. This means that biomarkers have potential as targets of interventions (Perera & Mooney, 1993).

Within-person biological variability may or may not be time-dependent (Hulka & Margolin, 1992). Biological specimens collected from the same person at different times might show changes because the person has aged, has been exposed to mutagenic substances or has been exposed to a medical procedure. The actual biological matrix, e.g. white blood cells, will have changed in the intervening time. Some cells will have died, while others may have been generated.

Time-independent biological variability will also occur because the distribution of matter (e.g. xenobiotics) can vary across organs and cells. Each biological sample will not capture the same biological material (Hulka & Margolin, 1992).

Intraperson variability has important implications for sample size and power in transitional studies (Hulka & Margolin, 1992). This has been demonstrated in cytogenetic studies, such as that by Hirsch *et al.* (1984) who found that cell-to-cell variability in the frequency of sister chromatid exchanges from an individual person was greater

than the variability in the mean frequency of sister chromatid exchanges from person to person. To compensate for this intra-individual variability, the number of cells scored per person was increased. Interperson variability in biomarker measures is generally addressed in the same way as other types of variable measures in epidemiological studies. The exception may be that the variability of biomarkers between persons may reflect acquired or inherited susceptibility which can serve as an effect modifier in relation to exogenous exposures of interest (Hulka & Margolin, 1992). Susceptibility to environmental agents is likely to arise from complex interaction between activating and inactivating chemical metabolizing enzymes. For example, Tang *et al.* (1995) have estimated that the combination of an inherited deletion of a gene for the detoxifying enzyme glutathione-S-transferase (GSTM1) and high PAH-DNA adduct levels confers a 12-fold risk of lung cancer.

The ability to measure biomarkers at the molecular and genetic levels has resulted in the identification of a degree of interperson variability not previously imagined. The major sources of this variability need to be accounted for prior to the use of biomarkers in etiological or public health applications. Characterization-type transitional studies may involve determining the prevalence of particular alleles in specific racial subgroups, evaluating the correlation between genotypic and phenotypic assays, estimating the likelihood and impact of allele misclassification and evaluating potential induction effects. Additionally, transitional studies can evaluate the biological plausibility of gene–environment interactions observed in etiological studies (Rothman *et al.*, 1995).

Applied transitional studies

Applied transitional studies are those that assess the relationship between a marker and the event that it marks, namely exposure, disease or susceptibility. These studies are often conducted on healthy subjects, and generally, in most cases involving biomarkers of exposure or effect, the biomarker is treated as the outcome variable. However, with susceptibility biomarkers and in certain study designs, a biomarker of exposure or effect may be used as an independent variable (e.g. serum organochlorines and breast cancer; carcinogen DNA adducts and lung cancer). Applied transitional studies are generally cross-sectional or short-term longitudinal designs and are not capable, in and of themselves, of establishing or refuting a causal relationship between a given exposure and disease. They do, however, provide mechanistic insight and may yield useful information on relationships between biomarkers and the events they represent (Rothman *et al.*, 1995). Where case–control studies that link a biomarker with disease are conducted, inferences about causality can sometimes be made. Another type of applied transitional study that has been reported involves the use of biomarkers in intervention studies. These include interventions with smokers, workers and the treatment of high-risk populations (e.g. in treatment with antioxidants, biomarkers are being used to monitor the efficacy of the intervention) (Perera & Mooney, 1993). The studies are designed to test not only the presence of the biomarkers, but also the level of change with alteration of the risk factor. These studies are transitional in that they establish the ability of the marker to serve as an early or intermediate 'end-point' with which to monitor efficacy or compliance.

The objective of the studies is to determine one of the following relationships:

- exposure/marker
- marker/disease
- exposure/susceptibility marker (high disease risk or low disease risk).

Applied transitional studies can be used to assess the attributable proportion of a particular biomarker. Hence, they build on the mechanistic knowledge obtained in characterization-type transitional studies.

Exposure/marker

The assessment of the relationship between an exposure and a marker represents much of the previous effort termed molecular epidemiology. Exposures to carcinogens have been evaluated to determine changes in DNA (DNA adducts) or proteins (e.g. haemoglobin or albumin adducts), or in some cases cytogenetic changes (e.g. chromosomal aberrations, micronuclei). These have generally been cross-sectional in nature. It is critical in such studies to be rigorous not only in the assessment of the marker but also in the assessment of exposure and

control for confounding. This is illustrated in the work of Mayer *et al.* (1991) and Schulte *et al.* (1992), who demonstrated the value of putting appropriate resources into the exposure characterizations when validating the relationship between exposure to ethylene oxide and formation of haemoglobin adducts. Too often, emphasis is placed only on the assay and not on the measure of exposure.

In characterizing the relationship between a marker and exposure, it is important that all sources and rates of exposure are considered, since a marker generally represents the integration of these. Attention should also be given to the toxicokinetics and natural history of the marker in order to understand the appropriate sampling and specimen collection times. A marker with a short half-life will not be detectable in samples a long time after exposure has ended. Moreover, there is a need to assess the relationship between the marker and different regimens of exposure (continuous or intermittent).

Marker/disease

The relationship between a marker and disease is often the most frequently considered issue in assessing whether a marker is ready for use in an epidemiological study. Of interest is how well the marker predicts or represents disease. These types of validation studies are difficult to accomplish because of the temporal factor. To identify an early change—i.e. a change in pathogenesis or a change predictive of disease—generally requires a prospective study, although cross-sectional clinical studies of heavily exposed individuals and case–control studies can be used to great advantage. However, when not using a prospective design, care must be taken to avoid biased associations. This is often difficult, and hence prospective studies are the best approach for validation. However, prospective studies are expensive and time-consuming, and few are conducted. For example, despite the large number of studies on cytogenetic markers, there is still little consensus on their predictive value, since most of the studies have been cross-sectional and suffer from temporal ambiguity. Specifically, in epidemiological terms, predictive value means the percentage of those who test positive for a marker who actually develop the disease. To perform the appropriate prospective studies of sister chromatid exchanges would take a large population and a relatively long time. The best and possibly only example of such a study is the Nordic prospective study on the relationship between peripheral lymphocyte chromosome damage and cancer morbidity in occupational groups (Brøgger *et al.*, 1990). Ten laboratories in four Nordic countries participated in a study of a combined cohort of persons (mostly from occupational groups) who had been cytogenetically tested. The cohort will be followed prospectively for cancer morbidity. The cohort comprises 3190 subjects, of whom 1986 (62%) have been scored for chromosome aberrations and 2024 (63%) have been scored for sister chromatid exchanges. Preliminary analysis indicates that chromosomal aberrations are associated with cancer.

Exposure/susceptibility to disease

Validated biological markers of susceptibility can serve as effect modifiers in epidemiological studies. Effect modification is a term with statistical and biological aspects. Statistically, effect modification is analysed by examining the joint effects of two or more factors. The interpretation of effect modification depends on the statistical method (e.g. multiplicative or additive) used to model interaction. From the biological perspective, effect modification can explain why two similarly exposed individuals do not develop a disease. The answer, in part, is individual variability in metabolic, detoxification and repair capabilities.

To validate a susceptibility marker, it is important to minimize misclassification, which can occur as a result of laboratory or epidemiological factors that affect phenotyping or genotyping (Rothman *et al.*, 1993). Next, it is necessary to demonstrate that the susceptibility marker either increases the biologically effective dose or elevates the risk of disease.

Genetic susceptibility markers can be characterized in terms of the six possible patterns of gene–environment interaction identified by Khoury *et al.* (1988). These are shown in Table 3. Khoury *et al.* (1988) recommend that an epidemiological approach be used to evaluate genetic marker–disease associations and their interaction with specific environmental risk factors. However, prior to conducting such studies, a major challenge is to determine which environmental factors might be involved in the etiology of a specific disease.

Table 3. Patterns of genotype–environment interactions observed in ecogenetic studies

Pattern	Effect of genotype in the absence of the environment	Specificity of environmental effect *vis-à-vis* genotype	Notations
1	Innocuous	Specific	$R_g = 1, R_e = 1$
2	Innocuous	Non-specific	$R_g = 1, R_e > 1$
3	Risk factor	Specific	$R_g > 1, R_e = 1$
4	Risk factor	Non-specific	$R_g > 1, R_e > 1$
5	Protective	Specific	$R_g < 1, R_e = 1$
6	Protective	Non-specific	$R_g < 1, R_e > 1$

Source: Khoury *et al.*, 1988.

In cancer epidemiology, interest has been focused on susceptibility genes that are common in the population and are generally considered to be polymorphisms (i.e. with a minor allele frequency of more than 1%), that are probably associated with relative risks under 10 (and as such do not exhibit familial patterns of inheritance) and that may interact with a particular exposure (Rothman *et al.*, 1995). The objective of applied transitional studies with regard to susceptibility markers is to determine whether the markers are effect modifiers. As mentioned earlier, the distinction between a transitional study and an etiological study becomes slightly blurred in this effort, so that such distinctions are arbitrary. The emphasis in a transitional study is to determine whether the marker can be used to investigate populations in terms of risk (i.e. serve as an outcome variable or as an independent variable).

Preparing for transitional studies

Selecting candidate markers

Many more markers are identified in laboratories than could reasonably be tested and used in the field. For this reason, there is a need to select among candidates which markers are beneficial for field testing and use. To gauge utility, it is helpful to envision a framework for candidate markers such as the continuum between exposure and disease. A potentially useful candidate marker will be one that can be related to some heuristic continuum and for which successful field testing will add relevant information to various etiological or public health questions. In some cases, a transitional study may only provide mechanistic information. This can be very useful but should be oriented towards confirming a hypothesized link in a continuum.

It is of great importance in identifying a continuum of events between exposures and disease that the marker represents a 'critical effect'. (Schulte, 1989; Borm, 1994). A critical effect is the biological marker that is deemed most representative of a particular component in the continuum and is ultimately most pathognomonic. This requires a series of independent studies, primarily toxicological, but also clinical and epidemiological. It is necessary to develop a hypothesis concerning the role of the marker in the development of the disease. As more causal components are identified, it becomes necessary to elucidate quantitative relationships of the kinetics, natural history and rates of transition along the continuum.

Determining level of effort

After a candidate marker has been identified, it is useful to determine what field testing is required. This is a useful exercise because it allows for research planning and funding and assures that a comprehensive approach is considered. The alternative is that a marker becomes labelled as validated and ready for field use when it is not ready or eligible, and this can lead to flawed and costly studies. An example of a check list for what needs to be considered is shown in Table 2.

Selecting candidate populations

A key factor in transitional studies is selecting and accessing populations with the kind of characteristics thought to be important to the testing of a

marker. These may be populations with contrasting levels of exposure, demographic or behavioural factors, or ethnicity; or they may be populations homogeneous for such factors. The similarity between these populations and those in whom the marker will ultimately be used for etiological or applied purposes is important.

Reporting results of transitional studies

Transitional studies can be highly informative regardless of their outcome. Those that show a positive relationship between a marker and exposure, disease or susceptibility are obviously important. However, those that show negative relationships between a biomarker and a particular event, or have small biomarker frequencies in a population subgroup, provide useful information about the utility, generalizability or limitations of a marker. Such negative results in well conducted studies should be published in the peer-reviewed scientific literature. Clearly, statistical power considerations should be discussed in such studies.

There appears to be a wide variation in the approaches used in reporting test and study results to participants in the research. Some investigators and their supporting organizations require all test and subject results to be communicated to participants regardless of clinical relevance and with the most truthful interpretation possible, while others limit reporting to those results that are clinically relevant. This difference hinges on the tension between beneficence and autonomy (see Schulte *et al.*, this volume, for a discussion of the ethical issues). Should participants be told only when something can be done and not told when there might be anxiety without benefit, or should participants have a right to information that is held about them regardless of whether it is considered by the holder to be of benefit? These questions need further consideration.

Support for transitional research

Transitional studies are important for the successful and effective use of biomarkers in cancer epidemiology. However, they are perceived as having neither the excitement nor the appeal of basic laboratory or etiological research or public health application. Thus, they are not widely and intensively supported by funding agencies. Since their outcome is the characterization of biomarkers and their limits, they are intermediate rather than end results about cancer causation or controls; yet without this effort, the end results may not be obtained. Currently, much of the work of transitional studies is subsumed in pilot or feasibility studies conducted prior to a larger study. This may not be the most effective way of conducting transitional research because researchers may be forced to trade off those funds available for conducting an etiological study against those available for the assessment of the utility and limits of a marker. More funding agencies should designate and support transitional biomarker studies if a wide range of useful tools are to become available.

References

Benichou, J. (1991) Methods of adjustment for estimating the attributable risk in case–control studies: a review. *Stat. Med.*, 10, 1753–1773

Bernard, A.M. (1995) Biokinetics and stability aspects of biomarkers: recommendations for application in population studies. *Toxicology*, 101, 65–71

Borm, P.J.A. (1994) Biological markers and occupational lung disease: mineral dust-induced respiratory disorders. *Exp. Lung Res.*, 20, 457–470

Brogger, A., Hagmar, L., Hansteen, I.L., Heim, S., Hogstedt, B., Knudsen, L., Lambert, B., Linnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerfving, S. & Sorsa, M. (1990) An inter-Nordic prospective study on cytogenetic endpoints and cancer risk. *Cancer Gene Cytogenet.*, 45, 85–92

Droz, P.O. (1993) Biological monitoring and pharmacokinetic modeling for the assessment of exposure. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 137–157

Fleiss, J.L. (1986) Statistical factors in early detection of health effects. In: Underhill, D.M. & Radford, E.D., eds, *New and Sensitive Indicators of Health Impacts of Environmental Agents*, Pittsburgh, PA, University of Pittsburgh Press, pp. 9–16

Hertzberg, V.S. & Russek-Cohen, E. (1993) Statistical methods in molecular epidemiology. In: Schulte, P.A. & Perera, F.P. eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 199–216

Hirsch, B., McGue, M. & Cervenka, J. (1984) Characterization of the distribution of sister chromatid exchange frequencies: implications for research design. *Hum. Genetics*, 65, 280–286

Hulka, B.S. (1990) Methodologic issues in molecular epi-

demiology. In: Hulka, B.S. Wilcoskey, T.C. & Griffith, J.D., eds, *Biological Markers in Epidemiology*, New York, Oxford University Press, pp. 214–226

Hulka, B.S. (1991) Epidemiologic studies using biological markers: issues for epidemiologists. *Cancer Epidemiol. Biomarkers Prev.*, 1, 13–19

Hulka, B.S. & Margolin, B.H. (1992) Methodological issues in epidemiologic studies using biologic markers. *Am. J. Epidemiol.*, 135, 200–209

Khoury, M.S., Adams, M.J., Jr & Flanders, W.D (1988) An epidemiologic approach to ecogenetics. *Am. J. Hum. Genetics*, 42, 89–95

Massey, B.S. (1986) *Measures in Science and Engineering* ,Chichester, Ellis Horwood

Mayer, J., Warburton, D., Jeffrey, A., Pero, R., Andrews, L., Walles, B., Toor, M., Latriano, L., Tang, D., Tsai, W.Y., Kuroda, M. & Perera, F.P. (1991) Biologic markers in ethylene oxide-exposed workers and controls. *Mutat. Res.*, 248, 163–176

Perera, F.P. & Weinstein, I.B. (1982) Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J. Chronic Dis.*, 35, 581–600

Perera, F.P. & Mooney, L.A. (1993) The role of molecular epidemiology in cancer prevention. In: DeVita, V.T., Hellman, S. & Rosenberg, S.A., eds, *Cancer Prevention*, Philadelphia, J.B. Lippincott, pp 1–15

Perera, F. & Santella, R. (1993) *Carcinogenesis*. In: Schulte, P.A. & Perera F., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 277–300

Rothman, N. (1993) Epilogue. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp 199–216

Rothman, N., Stewart, W.F., Caporaso, N.E. & Hayes, R.B. (1993) Misclassification of genetic susceptibility biomarkers: implications for case–control studies and cross-population comparisons. *Cancer Epidemiol. Biomarkers Prev.*, 2, 299–303

Rothman, N., Stewart, W.F. & Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev.*, 4, 301–311

Schatzkin, A., Freedman, L.S., Schiffman, M.H. & Dawsey, S.M. (1990) Commentary: validation of intermediate end points in cancer research. *J. Natl Cancer Inst.*, 82, 1746–1752

Schesselman, J.J. (1982) *Case-control Studies*, New York, Oxford University Press

Schulte, P.A. (1989) A conceptual framework for the validation and use of biological markers. *Env. Res.*, 48, 129–144

Schulte, P.A. (1992) The use of biological markers in occupational health research and practice. *J. Toxicol. Env. Health*, 40, 359–366

Schulte, P.A., Boeniger, M., Walker, J.T., Schober, S.E., Pereira, M.A., Gulati, D.K., Wojciechowski, J.P., Garza, A., Froelich, R., Strauss, G. et al. (1992) Biologic markers in hospital workers exposed to low levels of ethylene oxide. *Mutat. Res.* 278, 237–251

Schulte, P.A. & Perera, F.P. (1993) Validation. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 81–109

Schulte, P.A., Rothman, N. & Schottenfeld, D. (1993) Design consideration in molecular epidemiology. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 159–198

Shrout, P.E. & Fleiss, J.L. (1979) Intraclass correlations: uses in assessing rator reliability. *Psychol. Bull.*, 86, 420–428

Tang, D.L., Chiamprasert, S., Santella, R.M. & Perera, F.P. (1995) Molecular epidemiology of lung cancer: carcinogen-DNA adducts, GSTM1 and risk. *Proc. Am. Assoc. Cancer Res.*, 36, 284

Trock, B.J. (1995) Application of biological markers in cancer environmental epidemiology. *Toxicology*, 101, 93–98

Vineis, P., Schulte, P.A. & Vogt, R.F., Jr (1993) Technical variability in laboratory data. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press pp. 109–135

Wilcox, A.J. (1992) Molecular epidemiology: collision of two cultures. *Epidemiology*, 6, 561–562

Winn, D.M. & Gunter, E.W. (1993) Biologic specimen banks: a resource for molecular epidemiologic studies. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 217–235

Corresponding author
P.A. Schulte
National Institute for Occupational Safety and Health, 4676 Columbia Parkway, C-14, Cincinnati, OH 45226, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Logistics and design issues in the use of biological samples in observational epidemiology

J.D. Potter

Standard epidemiological study designs are well suited to answering questions involving the collection of biological samples. However, different designs are better suited—both for design and logistic reasons—to different questions. The strengths and weaknesses of each design are discussed in relation to markers of exposure, susceptibility and early outcome, and to markers used to classify cancers into biologically defined subsets.

Epidemiology has, for over a century, encompassed observational and experimental modes. The earliest modern epidemiological endeavour, which we all learn at our mentor's knee, is the story (usually, alas, oversimplified) of John Snow and the Broad Street pump. From very early on (but subsequent to the Snow experiment), epidemiologists have studied both reported and recorded data, on the one hand, and biological samples on the other. Infectious disease epidemiology, particularly acute outbreaks, requires both 'shoe-leather' epidemiology and the collection of specimens. Chronic disease epidemiology is similar, particularly where there are established biological markers that are central to the disease process: cholesterol in coronary heart disease; glucose and insulin levels in diabetes mellitus; uric acid in gout. What is important about both infectious diseases and the above-mentioned chronic diseases is that the clinical, experimental and epidemiological models evolved largely in concert with each other; biology, understood in one area, rapidly became incorporated into research endeavours in a different discipline.

Cancer is markedly different from other diseases in two ways. The first is that, unlike the infectious diseases and the above-mentioned chronic diseases, cancer is not regarded as a systemic disorder. Consequently, treatment is largely local (surgery, radiation), and there is little opportunity to understand the disease process when the treatment is successful; the later stages of disease and their systemic manifestations are regarded as signs of the failure of treatment. Secondly, in experimental models of cancer in animals, the cancer is induced by carcinogens (both chemical and viral) and promoters that are intended to model, but are essentially unrelated to, the human disease. Examples include the use of dimethylhydrazine (DMH) as a carcinogen in rat colon cancer, dimethylbenzanthracine (DMBA) as a carcinogen in mouse breast cancer, and phorbol esters as promoters in mouse skin models. As an aside, it is worth noting that animal 'models' of cancer were originally intended to mimic the end-point in order to study it better. However, we appear to have lost this perspective, and have made the error of assuming that because these models mimic the end-point, they also mimic the process.

Since the 1960s, as a result of studying human pathology of cancer, it has become clear that there are many early stages of cancer that are discernible and can be studied in their own right, e.g. cervical intra-epithelial neoplasia, adenomatous polyps of the colorectum, proliferative dysplasia of the breast and Barrett's oesophagus. More recently, it has also become clear that, even at a molecular level, specific carcinogens do not always act in the manner that was previously hypothesized (Jin *et al.*, 1996) and do not always act in humans in the same way as in animals (Kakiuchi *et al.*, 1995).

What is now beginning to emerge in the study of human carcinogenesis is the kind of convergence of clinical, experimental and epidemiological studies that has previously characterized the study of heart disease. Although, in the main, we lack the systemic markers and manifestations to be used as risk indicators of disease (such as lipoproteins, cholesterol and its subfractions, and hyper-

tension), there are, none the less, some probable systemic markers for certain cancers, and we do have an increasing understanding of the disease process at molecular level and of its manifestations at cell and tissue levels. Further, as a consequence of several decades of largely 'paper-and-pencil' occupational, environmental and dietary epidemiology, we have some fairly robust ideas about which exposures increase or decrease risk. Additionally, with developments in epidemiology and biostatistics, we know to what extent misclassification and measurement error attenuate our findings. Finally, extensive development in molecular and biochemical methods over the last two decades have improved markedly our capacity to detect and quantify components of relevant biological pathways in small samples collected on large numbers of people.

Against this background, it is appropriate to consider the methodological issues that arise regarding the incorporation of biological markers in epidemiological studies. This chapter covers the following: consideration of some of the logistics of sample collection in observational studies in the clinical setting and in population-based studies; the issues surrounding the use of markers of various stages in the cancer process—susceptibility, exposure, early biological effect, surrogate or intermediate end-points, and markers defining subsets of disease in each of these designs and settings; and the particular issue of studies to develop screening markers.

In undertaking the collection of human samples for epidemiological studies, the best design and the easiest logistics are sometimes in conflict. For most etiological questions, population-based case–control studies are preferable to those in which study subjects are recruited solely in the clinic/hospital population. For cohort studies, recruits from the general population who are in generally good health are preferable to patients (unless, of course, the cohort is focused on complications of the initial disease and its therapy).

By contrast, the recruitment of a hospital/clinic-based population makes the logistics of sample collection—susceptibility markers, exposure markers, early end-points or pathology samples—considerably easier. As with many epidemiological design and execution issues, it is sometimes necessary to establish some compromise between the methodologically desirable and the logistically feasible.

A variety of strategies exist. Almost all require collaboration across disciplines and a supportive, understanding study population. Matches between design and the question asked are outlined in the following sections. An attempt is made to identify the best way of answering specific questions. A different typology and approach have been proposed by Rothman *et al.* (1995).

Clinic-based cohort studies

Two different kinds of cohort studies can be established in the clinical setting. The first is a screening-based cohort where individuals are enrolled specifically following a negative screen (e.g. mammography, colonoscopy or Pap smear), and where specific biological samples, such as blood and benign lesions, are stored to establish the relationships between exposure, susceptibility, metabolic markers such as hormones or disease intermediate markers (e.g. adenoma or cervix pathology) and subsequent risk of disease. Examples of such studies include the New York Breast Cancer Study (Toniolo *et al.*, 1991) and the US National Polyp Study (Winawer *et al.*, 1993).

The second kind of cohort study that is well established in the clinic or hospital setting is the study designed specifically to follow individuals treated for disease with known pathology (including first primary cancers) in order to establish the risk of further disease (including second primary neoplasms). Here, this is termed a 'pathology' cohort. The biological markers that can be used in such studies include markers of susceptibility including specific genetic syndromes such as Li-Fraumeni, ataxia telangiectasia, hereditary non-polyposis colon cancer (HNPCC), the characteristics of the original tumour itself, and possibly biological measures of therapeutic exposures. One example of this design is the Children's Cancer Survivor Study (principal investigator Dr Leslie Robison) in the USA, which follows more than 20 000 survivors of childhood cancer who have lived at least 5 years after diagnosis and were disease-free at the time. Blood samples are being used to establish genetic susceptibility, and the cohort is to be followed to a variety of end-points so that the effects of genetic susceptibility, environmental risk factors and therapeutic exposures (radiation, chemotherapy) on subsequent disease can be ascertained.

Table 1. Optimality of design and relative ease of sample collection logistics in clinic-based cohort studies

Markers	Design	Logistics
Exposure (e.g. blood levels, adducts)	1–2[a]	1
Susceptibility (e.g. genes)	1–2	1
Intermediate end-point	2	2
Stratification of diseases into biologically defined subtypes	1	2
Biology of first tumour (pathology cohort) or pre-neoplastic pathology (screening cohort)	1	1

[a]1 = most optimal (design) or easiest (logistics).

Table 1 shows a ranking, on a scale of 1–3, for both design (where 1 is optimal) and logistics (where 1 is easiest) for the collection of a variety of materials in a clinic-based cohort study. In this type of study, the collection of baseline biological materials is clearly facilitated only if the recruitment of the study subjects actually occurs in the clinic. In the case of recruitment from a screening clinic, this is a relatively easy procedure: blood samples, etc., for markers of both exposure and susceptibility, can be collected, prepared and stored. This is also true for screening-associated pathological samples. Such material will usually be used in a nested case–control fashion as follow-up of the cohort reveals cases of the cancers of interest. For blood samples, where the focus is on, for example, specific exposure markers or high-prevalence genes such as metabolic enzyme polymorphisms, the follow-up of the cohort can investigate such markers in relation to a variety of cancer and pre-cancer outcomes. For low-prevalence inherited mutations and for specific intermediate pathological markers collected at screening recruitment (e.g. histology of a benign lesion), the follow-up will perhaps be restricted to just one cancer. Of course, the cohort can be followed up, using non-biological data, for any end-point.

As with any cohort, while internal comparisons will be valid, and interpretation of the strength of associations with biological markers will be clear, problems of generalizability may be more marked than is the case with a population-based cohort, given the selective nature of such clinic-based study populations (e.g. who comes for screening, or who gets the first primary cancer).

From a design standpoint, the timing of the collection of biological measures of exposure in relation to disease outcome is always problematic. If the period between collection and outcome is too long, there may be increasing misclassification of exposure levels; too short a period may result in early disease influencing the relevant exposure marker. Family-based cohort studies have similar logistic issues to population-based studies (Sellers *et al.*, 1995). The pathology cohort, in which the end-point may be a second primary cancer, in many cases will have biological material from the original diagnosis of cancer (or a benign lesion) and blood (though this is less likely) stored at the original institution. Retrieving this material and assembling such a cohort is a very time-intensive and labour-intensive procedure. The problem is exacerbated in the case of children where the original physician of record (probably a paediatric specialist) may no longer be in contact with the patient, where many of the potential study subjects will have moved and, with early adulthood, where a substantial proportion of the female subjects will have changed their surnames. Such cohorts can be assembled, but the collection of additional data and biological material (e.g. blood for exposure measures or genotyping) requires individual contact in the same manner as in a population-based cohort.

In this design, establishment of intermediate end-points requires active follow-up, as, with one or two exceptions (e.g. Pap registries, mammography registries), non-tumour end-points are not routinely registered; further, there may still be differential detection rates of such end-points within the cohort, depending on both the access to medical care and health behaviours of the study subjects. The etiology of intermediate end-points may be better studied in a case–control fashion. Original biological material (in both screening and pathology cohorts) can be used both to characterize the initial tumour and to establish (using normal tissue) the specific genotypes.

End-point pathological tissue can also be collected and cancer outcomes stratified into various

Table 2. Optimality of design and relative ease of sample collection logistics in clinic-based case–control studies

Markers	Design	Logistics
Exposure (e.g. blood levels, adducts)	1–3[a]	[1][b]
Susceptibility (e.g. genes)	1–2[c]	1
Intermediate end-point	1	1
Stratification of diseases into biologically defined subtypes	1	1

[a]1 = most optimal (design) or easiest (logistics).
[b]Appropriate for early disease/intermediate end-points, but not clinically detected cancer.
[c]The pattern of susceptibility genotypes may be distorted by the tendency of those with a family history to seek asymptomatic screening rather than presenting with early disease as a result of symptoms.

subtypes based on histopathology, gene expression, mutations, deletions, etc. No particular advantage is derived from the clinic-based cohort design for this purpose except the rather general one that this is likely to be a geographically restricted population, thus facilitating personal contact with patients and their physicians, which in turn allows active follow-up.

Clinic-based case–control studies

Table 2 shows, similarly to Table 1, the ranking for optimality of design and logistics when collecting various biological markers in the setting of clinic-based case–control studies. These studies are of particular value for investigating the etiology of precancerous lesions. For these diseases [e.g. early cervical intra-epithelial neoplasia (Brock *et al.*, 1988), colonic adenomatous or hyperplastic polyps, Barrett's oesophagus (Blount *et al.*, 1991; Meltzer *et al.*, 1994)], very few, if any, population-based registries exist and the detection of the disease, and therefore the identification of disease-free individuals to be enrolled as controls, is invasive. Such studies are only possible in the clinical setting. Further, the early stage of the disease reduces, but does not eliminate, the problem of the disease altering the biological markers/processes (as well as, of course, specific behaviours) of interest in the etiology of the disease. Such a setting is therefore the best approach to establishing the relation between exposures, whether measured by self-report or by biological markers, and the early stages of the neoplastic process. Similar benefit accrues to the study of specific genotypes (again both metabolic polymorphisms and specific germline mutations) and the known or suspected intermediate stages of the neoplastic process.

It is worth noting that, with a few interesting exceptions, the relationship between these intermediate disease markers and the cancers of which they are precursors is usually difficult to study, for both ethical and logistic reasons. For instance, in the case of established colonic adenomatous polyps, CIN III and benign proliferative disease of breast, the precursor lesions are removed and the natural history of the disease is disrupted—a particularly appropriate and satisfactory outcome for the patient but one which, none the less, removes the opportunity for studying a variety of pathophysiological and molecular processes.

One interesting exception is Barrett's oesophagus, where the progression of the dysplasia needs close monitoring but the malignant potential is not sufficient to justify the severe intervention of an oesophagectomy.

Clinic-based studies, where the etiology of cancer itself is the focus of study, allow the collection of end-point cancer tissue as well as blood. However, etiological questions using this material are confined to questions of the role of genotypes and etiology. Exposure markers are not appropriate in this setting, since the disease itself, the symptoms of the disease and its therapy may all alter any exposure marker or biological process thought to be a precursor of disease.

A particular additional use of this design is to establish screening markers. What is important in this variation of the study design is that biological markers are sought which classify people with and without a specific cancer or its precursors. Here, unlike studies of etiology, the temporal sequence of the biological markers and the outcome of interest is irrelevant if the association is consistent and allows a ready distinction between those with and those without the disease. The issue of study power needs to be borne in mind, but there are two facets

Table 3. Optimality of design and relative ease of sample collection logistics in population-based cohort studies

Markers	Design	Logistics
Exposure (e.g. blood levels, adducts)	1[a]	3
Susceptibility (e.g. genes)	1	3
Intermediate end-point	2	3
Stratification of diseases into biologically defined subtypes	1	2–3[b]

[a]1 = most optimal (design) or easiest (logistics).
[b]Disease stratification markers will vary in the degree of logistic complexity depending on whether the marker is routinely measured and recorded in a registry (e.g. estrogen receptors), whether it is measured and recorded in a chart (e.g. cytogenetics, FAB morphology, etc.), or whether it is measured for the study only.

to be considered. Although underpowered studies are in some danger of rejecting, as statistically non-significant, a promising marker that could be used in screening, it is nevertheless the case that for a screening marker to be of major use, it should readily (i.e. with small study size) distinguish between those with and those without the disease. The added complication is the possibility that two or more markers in combination may provide a useful screening battery. Input into the design of such studies should be obtained from biostatisticians, biology and clinical practice. Some further development of power calculations, methods and design for such studies seems warranted.

Population-based cohort studies

The population-based cohort study (see Table 3) has a number of advantages over other epidemiological study designs. Its prospective nature reduces bias, both in the collection of histories of exposure and in the use of biological measures of exposure. There are no specific advantages over case–control studies in the use of genotyping data nor in the capacity to stratify the outcomes of interest on biological subtypes.

A specific advantage does accrue, however, if, using regular re-contact of cohort members, it is possible to establish aspects of the natural history of disease via exploration of specific intermediate markers and processes. However, unless the cohort is routinely screened, there may be misclassification in the determination of such intermediate end-points. One possibility, of course, is to restrict comparisons to those who have been screened as a matter of routine care. The Harvard group have used such an approach to the study of adenomatous polyps within their cohort studies (Giovannucci *et al.*, 1993).

Subsequent exposure-marker collection is also facilitated. This may be relevant when it is intended to show that cumulative exposure, changes in exposure level or changes in metabolic state are important in the progression to cancer. This has not been used much in cohort studies of cancer, although it has been used, for example, in cohorts of lead-exposed children. Some further biostatistical methodological development to optimize the use of sequential exposure measures is warranted. As with the clinic-based cohort studies, the problem of the truncation of the disease process by the removal of early stages of the neoplastic process remains.

A possible disadvantage of the cohort design, as compared with the case–control design, is the fact that, usually, cohorts are somewhat restricted in their age range and thus may not provide the full picture if cancers at specific ages are determined differently, i.e. if there is an interaction between exposure markers and age, or between metabolic markers and age.

The logistics of sample collection in a population-based cohort study are the most problematic, as most cohorts are not able to be recruited via a clinical facility. Whether recruited by mail or by a specially established clinic, costs of biological collection are high. As with the clinic-based cohort study, blood samples or buccal cells can be used to establish a wide variety of genotypes from very little material using PCR, but studies of serum or red-cell markers of exposure (e.g. serum nutrient levels, haemoglobin adducts) will rapidly deplete the stored samples. Nested case–control analysis remains the design of choice—as is planned for the EPIC study (Riboli 1992).

Population-based case–control studies

This design is inferior to the cohort design in the biological measurement of exposure and, of course, in the far more limited focus of the outcome

Table 4. Optimality of design and relative ease of sample collection logistics in population-based case–control studies

Markers	Design	Logistics
Exposure (e.g. blood levels, adducts)	No[a,b]	3
Susceptibility (e.g. genes)	1	3
Intermediate end-point	[1]	No[c]
Stratification of diseases into biologically defined subtypes	1	2–3[d]

[a]1 = most optimal (design) or easiest (logistics).
[b]It may be acceptable for studies of very early stage disease (T_1; M_0; N_0), but the problem of disease affecting the biological marker of exposure precludes more general use.
[c]With a few exceptions, e.g. Pap smear registries or accurately self-reported intermediate lesions such as pigmented naevi, there will be few opportunities to conduct population-based case–control studies on intermediate end-points.
[d]See footnote 'b' to Table 3.

measures (single rather than multiple diseases). Its major advantages (Table 4) are its lower cost (although the cohort design almost certainly involves a lower cost per cancer studied) and its broader age range. The best use of the population-based case–control study is in the exploration of the role of specific genotypes in etiology (Heckbert *et al.*, 1992; Kadlubar *et al.*, 1992) (again, both the rarer genetic mutations and the more common metabolic polymorphisms may be studied) and in the end-point stratification by molecular, histological or other means. The number of samples needed to answer either of these questions is markedly lower than in the cohort design, but the logistics of blood collection is likely to be just as complicated as in the cohort, since neither cases nor controls are being recruited via a clinical route. Exposure markers (for design reasons) and intermediate markers (for logistic reasons) frequently cannot be studied in a useful fashion in this design. The collection of exposure markers in a case–control fashion is not likely to be useful, as noted above in the discussion on clinic-based case–control studies. With a few exceptions (e.g. population-based Pap smear registries or well-measured self-reported intermediate markers such as pigmented naevi), there are no opportunities to conduct population-based case–control studies of intermediate markers.

In summary, each of the standard studies used in observational epidemiology acquires specific advantages and disadvantages when biological material is used to measure exposure or susceptibility, or is obtained in order to define specific intermediate end-points or to establish subsets of cancer outcomes. The screening clinic-based cohort study is most valuable in the follow-up to a specific cancer end-point in individuals known to be free of the disease at baseline. The focus is on both exposure and susceptibility markers relevant to that end-point. The clinic-based case–control study is of particular value in establishing the etiology of pre-cancerous lesions and in developing promising screening markers. The population-based cohort study gives the best opportunity for a generalizable study of etiology with exposure markers analysed in nested case–control fashion. The population-based case–control study provides the most cost-effective way of studying genetic susceptibility in the etiology of specific cancers.

References

Blount, P.L., Ramel, S., Raskind, W.H., Haggitt, R.C., Sanchez, C.A., Dean, P.J., Rabinovitch, P.S. & Reid, B.J. (1991) 17p allelic deletions and p53 protein overexpression in Barrett's adenocarcinoma. *Cancer Res.*, 51, 5482–5486

Brock, K.E., Berry, G., Mock, P.A., MacLennan, R., Truswell, A.S. & Brinton, L.A. (1988) Nutrients in diet and plasma and risk of *in situ* cervical cancer. *J. Natl Cancer Inst.*, 80, 580–585

Giovannucci, E., Stampfer, M.J., Colditz, G.A., Rimm, E.B., Trichopoulos, D., Rosner, B.A., Speizer, F.E. & Willett, W.C. (1993) Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J. Natl Cancer Inst.*, 85, 875–884

Heckbert, S.R., Weiss, N.S., Hornung, S.K., Eaton, D.L. & Motulsky, A.G. (1992) Glutathione S-transferase and epoxide hydrolase activity in human leukocytes in relation to risk of lung cancer and other smoking-related cancers. *J. Natl Cancer Inst.*, 84, 414–422

Jin, Z., Houle, B., Mikeev, A.M., Cha, R.S. & Zarbl. H. (1996) . Alterations in Hras1 promoter conformation during NMU-induced mammary carcinogenesis and

pregnancy. *Cancer Res.*, 56, 4927–4935

Kadlubar, F.F., Butler, M.A., Kaderlik, K., Chou, J. & Lang, N.P. (1992) Polymorphisms for aromatic amine metabolism in humans: relevance for human carcinogenesis. *Environ. Health Persp.*, 98, 69–74

Kakiuchi, H., Watanabe, M., Ushijima, T., Toyota, M., Imai, K., Weisburger, J.H., Sugimura, T. & Nagao, M. (1995) Specific 5′-GGGA-3′ → 5′-GGA-3′ mutation of the *Apc* gene in rat colon tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine. *Proc. Natl Acad. Sci. USA*, 92, 910–914

Meltzer, S.J., Yin, J., Manin, B., Rhyu M.-G, Cottrell, J., Hudson, E., Redd, J.L., Krasna, M.J., Abraham, J.M. & Reid, B.J. (1994) Microsatellite instability occurs frequently and in both diploid and aneuploid cell population of Barrett's-associated esophageal adenocarcinomas. *Cancer Res.*, 54, 3379–3382

Riboli, E. (1992) Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann. Oncol.*, 3, 783–791

Rothman, N., Stewart, W.F. & Schulte, P. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev.*, 4, 301–311

Sellers, T.A., Anderson, V.E., Potter, J.D., Bartow, S.A., Chen, P.-L, Everson, L., King, R.A., Kuni, C.C., Kushi, L.H., McGovern, P.G., Rich, S.S., Whitbeck, J. & Wiesner, G. (1995) Epidemiologic and genetic follow-up study of 544 breast cancer families: design and methods. *Genetic Epidemiol.*, 12, 417–429

Toniolo, P.G., Pasternak, B.S., Shore, R.E., Sonnenschein, E., Koenig, K.L., Rosenberg, C., Strax, P. & Strax, S. (1991) Endogenous hormones and breast cancer: a prospective cohort study. *Breast Cancer Res. Treat.*, 18, S23–S26

Winawer, S.J., Zauber, A.G., Ho, M.H., O'Brien, M.J., Gottlieb, L.S., Sternberg, S.S., Wayne, J.D., Schapiro, M., Bond, J.H., Panish, J.F., Ackroyd, F., Shike, M., Kurtz, R.C., Hornsby-Lewis, L., Gerdes, H., Stewart, E.T. & National Polyp Study Workgroup. (1993) Prevention of colorectal cancer by colonoscopic polypectomy. *New Engl. J. Med.*, 329, 1977–1981

J.D. Potter
Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Methodological issues in the use of biological markers in cancer epidemiology: cohort studies

D.J. Hunter

In this chapter we summarize the major strengths and weaknesses of cohort studies; consider how these characteristics influence the use of biomarkers in cohort studies; briefly review considerations of statistical power, design and the influence of measurement error in cohort studies; and discuss some of the emerging ethical considerations that relate to the use of biomarkers in prospective studies.

Most of the considerations that make prospective studies an attractive study design apply to exposures in general, not only to biomarkers.

Strengths

Exposure measured before outcome

The major strength of cohort studies is that the temporal relationship between exposure and outcome reflects the sequence of the causal pathway, i.e. exposure is measured before disease rather than afterwards, as in case–control (retrospective) studies. Thus, we can be assured that the biomarker level in cases preceded the outcome. This timing has three important implications:

1. Provided that undiagnosed or pre-clinical disease has not altered the biomarker level (see below), we can be sure that the biomarker has not been influenced by the disease or its treatment.
2. In a long-running cohort, a range of different times will exist between biomarker collection and diagnosis of the outcome. This can be used to investigate hypotheses relating to different latent periods between exposure and outcome, although the timing of first exposure (if it occurs prior to baseline) is usually uncertain.
3. Other exposures that are assessed prior to disease, such as questionnaire measures of diet, exercise or other lifestyle characteristics, should also be free of the influence of diagnosis of disease, and biomarker–environment interaction analyses should not be biased by the differential recall of cases and controls (recall bias) which may occur in case–control studies.

Selection of controls

For results to be unbiased in a case–control study, the controls must be representative of the population that gave rise to the cases (Breslow & Day, 1980); deviations from this principle may lead to selection bias. The choice of a control series that meets this criterion is frequently difficult. Random sampling from the underlying population may be impractical, even if this population can be defined; the source population for cases presenting at a cancer referral hospital, for instance, may be very difficult to specify. In a cohort study, the source population that gave rise to the cases is explicitly defined (cohort members at risk of disease), and thus selection bias should be minimal as long as follow-up rates are high. As it is rarely feasible to measure biomarker levels of all cohort members at baseline, a sample of cohort members is usually tested, using either a nested case–control or case–cohort design (see below).

Participation

Selection bias can also be introduced if less than 100% of cases and eligible controls choose to participate. Participation rates, already low in some settings, such as large cities in the USA, may be further adversely affected in studies incorporating biomarkers, as collecting blood, urine or other samples usually requires more effort from participants than studies using questionnaire-based measures alone. In a cohort study, participation theoretically should be 100% if reserves of specimen are available for all subjects and follow-up is complete.

Weaknesses

Statistical power

The major weakness of cohort studies is that the number of cases of even the more common diseases is limited unless follow-up time is very long. A large cohort may not produce enough cases of uncommon diseases even if follow-up is maintained until the last cohort member dies. This may be particularly troublesome for studies using biomarkers, as only a subset of people willing to be in a questionnaire-based cohort study may be willing to undergo the additional inconvenience associated with providing a specimen such as blood or urine for biomarker analysis. Solutions to this problem include pooling information from multiple cohorts [for example, Mueller *et al.* (1989) obtained sera from cases and controls in several longitudinal studies to study Epstein–Barr virus antibody profiles and risk of developing Hodgkin's lymphoma], conducting a study in a high-risk cohort (e.g. uranium mine workers for lung cancer) or conducting a case–control study, in which the population giving rise to the cases is much larger than is feasible for a typical cohort study.

Specimen size and appropriateness

Some biomarkers require large amounts of specimen (e.g. several hundred millilitres of blood for certain pesticide residues), or unusual specimens that it may be impossible to collect in a prospective manner from large numbers of subjects (e.g. adipose biopsies).

Level of detail of other exposures

In general, the level of detail for most questionnaire-based exposures is lower in a prospective study than in a retrospective study for which the questions can be focused on the subset of exposures of direct interest with respect to the outcome under investigation. Furthermore, unless exposure history is updated frequently (many cohorts measure exposures only once at baseline due to the cost of obtaining updated measurements) then the exposure history between the baseline questionnaire or interview and disease occurrence will be missing in a cohort study. Obtaining follow-up measurements to supplement baseline measurements may be particularly problematic for biomarkers, because the expense of collecting and storing specimens is substantial.

In summary, for relatively common outcomes for which sufficient cases are available, then the prospective design of a cohort study utilizing biomarkers has substantial advantages over retrospective designs, although detailed questionnaire-based or interview-based information on other exposures may be lacking.

Timing of specimen collection and disease

In any cohort study there will be a range of times between the date of specimen collection and the date of disease, due to staggered dates of collection, as well as a variation in the onset of disease.

The influence of pre-clinical disease

A major concern in cohort studies of short duration is the possibility that the disease process has influenced the biomarker level, even if disease is undiagnosed at time of specimen collection. This problem is most likely to occur if the average follow-up time is months or 1–2 years, and can be minimized if procedures for disease screening at baseline are used—for instance, if the end-point is lung cancer, through the exclusion of persons with recent weight loss or lesions on chest X-ray. Alternatively, excluding cases occurring in the first 1–2 years of follow-up may eliminate this effect of pre-clinical disease. However, distinguishing between this source of bias and a genuine association of a biomarker with disease that is only manifested early in follow-up (e.g. a late promotion effect in cancer) can be difficult. The long-running controversy about whether low serum cholesterol levels are associated with increased risk of cancer is a good example. In several prospective studies, even those such as the Multiple Risk Factor Intervention Trial (MRFIT) in which a physical examination was conducted at baseline, persons with lower serum cholesterol levels were at higher risk of cancer diagnosis; this effect was limited to the first 2 years of follow-up (Sherwin *et al.*, 1987). Although the consensus (Lewis & Tikkanen, 1994) is that this is an example of bias introduced by pre-clinical disease, there are still those who argue that it is a valid and potentially causal association.

Inference from a single biomarker level

A major weakness of cohort studies in general is that exposure data may be sparse and only avail-

Table 1. Intraclass correlation coefficients (ICC) for the repeated measurement of plasma hormone levels and the resulting observed relative risk for specified true relative risks of 1.5–2.5

		Observed relative risk given specified true relative risk		
Plasma hormone	**ICC**	**1.5[a]**	**2.0**	**2.5**
SHBG (nmol/l)	0.92	1.5[b]	1.9	2.3
Percentage of free estradiol	0.80	1.4	1.7	2.1
Estradiol (pg/ml)	0.68	1.3	1.6	1.9
Prolactin (ng/ml)	0.53	1.2	1.4	1.6

[a]True relative risk.
[b]Observed relative risk.
From Hankinson *et al.* (1995).

able for a single point in time (e.g. baseline in a prospective study which does not update exposure). This may have several consequences, as described below.

Misclassification due to random within-person variation. Within-person variation is a characteristic of most biomarkers that are not fixed characteristics, such as genotype. Substantial variation may be minute-to-minute (e.g. blood noradrenaline), hour-to-hour (e.g. blood glucose), day-to-day (e.g. 24-hour urine sodium excretion), month-to-month (e.g. plasma β-carotene), or longer, and may have a diurnal rhythm (e.g. plasma cortisol), may be related to the menstrual cycle (e.g. plasma hormones), or may have a seasonal basis (e.g. plasma vitamin D levels). If this variability is random and non-differential between future cases and controls, then the consequence will be bias to the null in the relative risk calculated from the biomarker. This attenuation of relative risks is proportional to the amount of within-person variability, and can be substantial. For example, Hankinson *et al.* (1995) calculated the intraclass correlations between three measurements of levels of four hormones among 79 postmenopausal women; these ranged from 0.53 (prolactin) to 0.92 (sex hormone binding globulin, SHBG). For a true relative risk of 2.5, the observed relative risk would be 2.3 for SHBG, but only 1.6 for prolactin, if the relative risk was based on a single measurement rather than the average of three (Table 1).

Time-integration. In most studies of the epidemiology of chronic diseases, the exposures of interest are long-term average exposures, because the induction periods for diseases such as cancer or cardiovascular disease are usually thought to be of the order of years or decades. Thus, it is usually desirable that a biomarker reflect cumulative exposure over at least months or years. Time-integration is often a function of the nature of the biological sample being assayed. For many nutrients, for instance, concentrations in erythrocytes are less susceptible to short-term fluctuations than in plasma or serum (Hunter, 1990). Concentrations in adipose tissue frequently represent a longer exposure history than those in erythrocytes. A 24-hour urine sample is more likely to be representative of long-term intake than is a random urine sample. Feasibility constraints are frequently paramount in determining the choice of biological specimen to be obtained in a cohort study; however, consideration should always be given to the implications for the time-integration of the biomarkers of interest.

Inference from multiple biomarker levels. An obvious method for increasing the time-integration of exposure is to obtain specimens at several points in time. These can be averaged, or persons with sustained high levels can be compared with those with sustained low levels. Methods have been given to calculate the number of replicate measurements required to estimate the 'true' underlying mean value of a biomarker within a specified

range of error (Lui *et al.*, 1979), if within-person variation is assumed to be random. If within-person variation is not random, but is due to changes in behaviour or secular trends in exposure, then multiple measurements may be useful to account for this source of exposure misclassification.

The obvious problem in a large cohort study is that obtaining samples at more than one time greatly increases the expense of sample collection and storage, and the burden on study subjects.

Inference from biomarker/disease associations. The strength of the underlying association between a biomarker level and disease depends on whether the biomarker is measured at the point in the causal pathway when this association is maximal, or whether it is measured at some other point. Conventional theories of cancer causation distinguish between early events in the causal pathway ('initiators') and later events ('promoters'). It is increasingly clear that the latency period between exposure and cancer diagnosis can be very long, e.g. three decades or more for the deleterious effect of smoking on colorectal cancer (Giovannucci *et al.*, 1994), compared with about one decade for the protective effect of aspirin on this cancer (Giovannucci *et al.*, 1995). Measurement of a biomarker of a putative initiating exposure may thus need to be several decades prior to disease; measurement of more proximal exposures, such as a late promoter, could be several years prior to exposure. If the measurement is made at the wrong time, however, the association may be attenuated or missed altogether.

The within-person variability of the biomarker is also relevant to determining how close to the time of maximum effect the biomarker needs to be measured. At the extreme, if the within-person variability is nil, e.g. a genotype marker, then the time of measurement prior to disease is not critical. In contrast, if the correlation of levels of a continuous biomarker over time is low, then time of sampling will need to be close to the time of maximum effect, if this effect is to be estimated accurately.

Measurement of within-person variability

If within-person variability in biomarker levels is random, then knowledge of the correlation in a population of a single measurement with the average of multiple measurements can be used to 'correct' for attenuation of relative risk estimates introduced by the fact that a single measure non-differentially misclassifies subjects with respect to their true average exposure. Even a single repetition of a measurement can be used to estimate the correction factor; in fact, it has been shown (Willett, 1990) that the correction factor is more precisely estimated in data with two measurements on a large number of people than with multiple measurements on a smaller number. To be generalizable to the larger cohort, the sample asked to participate in a reproducibility study would ideally be a random sample of the cohort. This is rarely feasible, particularly in geographically dispersed cohorts, and is usually further limited by less than perfect participation. Care should therefore be taken in interpreting reproducibility study information from less representative populations.

Specimen collection issues

Collection and storage of a large specimen bank is a complex operation with numerous pitfalls.

A consideration that is particularly important in prospective studies is how to minimize the loss of information due to exhaustion of the sample. In case–control studies this may be less problematic, as much more sample may be available than it is feasible to collect and store from the large numbers of participants enrolled in a typical cohort study. The problem of sample exhaustion is obviously most acute with respect to incident cases of disease; if every biomarker of potential interest is entertained, then a few millilitres of plasma can vanish very quickly. Due to the advantages of cohort data, interest from the scientific community may be great, and requests for samples frequent. Refusal to part with precious samples may be met with hostility from those who are refused. It may be wise for those in charge of specimen banks to consider a number of common-sense procedures including:

- collaborating with laboratory colleagues to minimize the amount of sample required to perform an assay with acceptable precision; for example, Laden *et al.* (1997) collaborated with Dr Mary Wolff of Mt Sinai Medical Center on the development of methods to reduce the amount of plasma required to perform assays of DDE and PCBs from 1 ml to 0.5 ml with equivalent precision;

- seeking collaborators who can measure a range of biomarkers in simultaneous multiplex fashion on one aliquot of sample;
- organizing an advisory board that can review requests for samples, establish priorities and share responsibility for decisions on the purposes for which samples are to be used.

Although these concerns are particularly acute for case samples, it is also desirable not to exhaust samples from participants selected as controls, as controls may become future cases. For nested case–control analyses of multiple biomarkers, it is usually most efficient to choose a single set of controls, in order to ensure comparability of relative risks associated with each biomarker, as well as to minimize the demands of aliquoting and sample handling. This may rapidly exhaust the sample from these controls however, and thus many investigators establish rules that mandate that a new control is selected when the amount of sample for a control falls below some minimum value, in order to preserve specimens for future analyses if the control becomes a case.

Matching controls to cases

Matching of cases to controls on a variety of characteristics may be used by investigators to increase the efficiency of statistical analyses. In biomarker studies, an additional level of matching is usually required to ensure the comparability of biomarker information between cases and controls.

Sample comparability

Biomarker levels are not fixed attributes of specimens, but may be altered by a wide range of procedures inherent in sample collection and storage. Thus, the overriding requirement for validity in biomarker studies is that case and control specimens are handled in the same way. In prospective studies, an additional factor to consider is length of storage; levels of many biomarkers degrade over time, and even under ideal storage conditions this degradation may be substantial (Bolelli *et al.*, 1995). Thus, it is usually necessary to match cases and controls on duration of storage, in addition to other factors such as season of specimen collection (for biomarkers with seasonal variation), number of freeze–thaw cycles to which the specimen has been exposed, whether the specimen was collected fasting or non-fasting, etc. Finally, batch-to-batch variability in laboratory assays can be substantial; thus, where possible, cases and controls should be run together in the same laboratory batches for maximum comparability, whatever the study design.

Selection of controls: nested case–control or case–cohort

In a large cohort study it is rarely desirable to measure levels of a biomarker on all cohort members at baseline, largely due to the expense this would entail. Some exceptions may exist—for example, assays that require fresh specimen or can be automated, e.g. haemoglobin or white cell count. The usual approach, however, is to select all the incident cases, and to sample the control information.

Nested case–control

In a nested case–control study, the controls are selected from the population which gave rise to the cases. Ideally, controls are selected for each case from the risk set of participants eligible to become a case at the time the case was diagnosed. It may be computationally easier to select from the set of participants at baseline who did not become a case during follow-up. If the amount of specimen needed is large, this latter procedure may also be preferable as it means that all incident cases are analysed; in the former risk set sampling, some case specimens may be unavailable if the case had been chosen as a control for a case occurring earlier in follow-up. It has been shown that if the cumulative incidence of disease is low (e.g. less than 5%), sampling from the baseline cohort introduces negligible bias (Langholz & Thomas, 1990).

Case–cohort

One of the inconveniences of nested case–control studies is that as the number of case–control sets increases, the laboratory work required to conduct analyses increases. A potential solution to this problem is the case–cohort design, in which the population distribution of the biomarker of interest is measured at baseline in a subcohort, and then information about cases is added subsequently as they occur. The calculations necessary to compute the variance of relative risks in this study design are complex (Prentice, 1986), but are now available in statistical packages (Hirosoft

International Corporation, 1993) which make their calculation convenient.

The major problems with the case-cohort design in biomarker studies are as follows:

1. If sample degradation over time occurs, differential degradation will occur in case and subcohort specimens.
2. If laboratory drift in the biochemical analysis is a concern, then having all the subcohort samples analysed at a different point in time from the cases may introduce substantial bias. This problem tends to be less acute when the biomarker is a qualitative trait such as genotype.
3. Laboratory personnel are more easily unblinded to case and non-case specimens in this design.

These problems make the case–cohort design of limited utility in prospective biomarker studies.

Choosing the cohort

The range of exposures in a cohort of members of the general population will reflect the exposures typical in the general population, and rare exposures or unusually low or high exposure levels may be rare or even absent. If the research depends on having large numbers of persons at the extremes of exposure, or persons with rare exposures, e.g. certain occupational exposures, then oversampling these persons or restricting to certain groups may be desirable.

Ethical issues

If a biomarker is known to be predictive of disease probability, or if our own research shows that this is the case, then the question of whether to inform the participants in the study of their biomarker result should be asked. This issue is particularly germane in prospective studies in which repeated re-contact with cohort members may occur during follow-up. In general, communicating this information to cohort members has not been the norm in epidemiological studies for a number of reasons, including the following:

- relative risks of the order of 1.1–3.0 are considered sufficiently low that the information is not sufficiently predictive to be clinically useful;
- the biomarker level is not considered definitely predictive of disease and more studies are needed;
- no interventions are established that could help a person at higher risk to avoid disease;
- if a screening method is available (e.g. mammography, sigmoidoscopy), the recommendations on frequency of screening are similar, whether or not the information on biomarker level is known;
- it would be logistically difficult or impossible to communicate the information to cohort members;
- the biomarker is clinically available and routinely performed (e.g. serum cholesterol);
- the within-person variability of the biomarker is so high that a value from the past (e.g. cohort baseline) would be poorly correlated with the current value;
- information (for cases) about pre-diagnostic levels has no prognostic significance, and thus is not useful for someone who has already been diagnosed with disease.

Another ethical norm in cohort studies is that explicit consent to the exact biomarker being measured has not been mandatory, as this may vary depending what disease the case developed, and with the rapid developments in the field of potential biomarkers, it was impossible to predict at the time of enrolment of subjects which biomarkers might be of scientific importance in the future. Thus, a global consent to research use of the biological specimen has usually been assumed to be adequate for the conduct of biomarker studies.

These norms are challenged by the rapid growth of potential markers of genetic susceptibility. High penetrance susceptibility genes (relative risks may be 50 or higher) are usually identified through family studies, in which the occurrence of a particular disease leaves no one in the family in doubt of what disease is being studied; in which a powerful incentive may exist for family members to participate in the research; in which issues of clinical treatment, screening and prophylactic procedures (e.g. prophylactic mastectomy in breast cancer kindreds) are substantial; and in which relatively few participants are involved and thus relatively intensive counselling is possible and appropriate. In addition, because these studies involve family members, and frequently children, major issues of confidentiality and/or coercion may arise, and thus procedures for informed consent must be rig-

orous. For these same reasons, counselling to warn participants of the potential psychological impact of knowledge of inherited susceptibility, and post-test counselling to mitigate potential harm, is frequently intensive.

Biomarkers of lower degrees of risk may be established through the 'candidate gene' approach: the frequency of polymorphisms that are known to alter the function of genes which might plausibly be involved in cancer etiology may be compared in cases and controls. This may result in the identification of biomarkers which convey much lower degrees of risk (e.g. relative risks of 1.5–3.0) or which only convey risk in the presence of certain environmental exposures (if a gene– environment interaction exists).

For all of the reasons specified above, prospective studies of both low and high penetrance genes are preferable to retrospective studies, in order to provide the most unbiased information. This information may be important in advising persons with these genotypes on their age-specific probability of disease and on lifestyle modifications they could make to reduce this probability, and may have a bearing on whether general population screening recommendations are adequate in their situation.

However, because genetic screening is involved, some experts hold that ethical norms previously thought to be adequate for the assumption of consent in prospective studies involving biomarkers are inadequate. A recent statement by members of the Ethical, Legal, Scientific, Implications (ELSI) panel of the US National Institute of Human Genome Research (Clayton *et al.*, 1995) points out the potential for harm to study participants if confidentiality is breached in genetic susceptibility research, and underlines the lack of consensus on appropriate standards of informed consent in this area. The extreme position is that any biomarker of genetic susceptibility requires the same explicit and detailed informed consent that is the norm in family studies, no matter what the degree of risk associated with this biomarker, and no matter what the state of the science is about this risk. Adoption of this extreme would render many epidemiological studies unfeasible, in particular most cohort studies, for the following reasons:

- It is difficult to predict who will be tested, and for what, at the start of a study which will be analysed in a nested case–control design. Covering all possibilities would provide participants with an overwhelming amount of information and potential choices, and would probably corrupt rather than enhance the process of informed consent.
- New biomarkers are becoming available that may not have been imagined at the start of the study, and thus prospective consent is impossible.
- Reconsenting participants for every new marker would involve multiple re-contacts, and participation rates would almost certainly fall, even in studies in which re-contact is possible. Many prospective studies, however, are designed to take advantage of passive follow-up through cancer registries, and re-contact may not be feasible.

If epidemiology is to make its appropriate, and socially desirable, contribution to these studies of biomarkers of genetic susceptibility, then guidelines need to be established which are appropriately respectful of the rights and concerns of subjects, but which do not rule out prospective studies of this whole class of biomarker research. These guidelines should be drafted by epidemiologists, ethicists, lawyers, clinicians, activists and representatives of research subjects, working together to resolve these issues.

References

Bolelli, F., Nuti, P., Micheli, A., Sciajno, R., Franceschetti, F., Krogh, V., Pisani, P. & Berrino, F. (1995) Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma. *Cancer Epidemiol. Biomarkers Prev.*, 4, 509–513

Breslow, N.E. & Day, N.E. (1980) *Statistical Methods in Cancer Research*, Vol. 1, *The Analysis of Case–Control Studies* (IARC Scientific Publications No. 32), Lyon, International Agency for Research on Cancer,

Clayton, E.W., Steinberg, K.K., Khoury, M.J., Thomson, E., Andrews, L., Kahn, M.E., Kopelman, L.M. & Weiss, J.O. (1995) Informed consent for genetic research on stored tissue samples. *J. Am. Med. Assoc.*, 274, 1786–1792

Giovannucci, E.L., Colditz, G.A., Stampfer, M.J., Hunter, D.J., Rosner, B.A., Willett, W.C. & Speizer, F.E. (1994) A prospective study of cigarette smoking and risk of colorectal adenoma and cancer in women. *J. Natl Cancer Inst.*, 86, 192–199

Giovannucci, E., Egan, K.M., Hunter, D.J., Stampfer, M.J., Colditz, G.A., Willett, W.C. & Speizer, F.E. (1995) Aspirin

and the risk of colorectal cancer in women. *N. Engl. J. Med.*, 333, 609–614

Hankinson, S.E., Manson, J.E., Spiegelman, D., Willett, W.C., Longcope, C. & Speizer, F.E. (1995) Reproducibility of plasma hormone levels in postmenopausal women over a 2–3 year period. *Cancer Epidemiol. Biomarkers Prev.*, 4, 649–654

Hirosoft International Corporation: EPICURE/PEANUTS Software (1993) *The PEANUTS Program. EPICURE User's Guide*. Seattle, WA, Hirosoft

Hunter, D.J. (1990) Biochemical markers of dietary intake. In: Willett, W., ed., *Nutritional Epidemiology*. New York, Oxford University Press, pp. 143–216

Laden, F., Wolff, M.S., Niguidula, N.J., Spiegelman, D., Hankinson, S.E. & Hunter, D.J. (1997) Reduced aliquot size for a plasma organochlorine assay for use in epidemiologic studies. *Cancer Epidemiol. Biomarkers Prev.* (in press)

Langholz, B. & Thomas, D.C. (1990) Nested case–control and case–cohort methods of sampling from a cohort: a critical comparison. *Am. J. Epidemiol.*, 131, 169–176

Lewis, B. & Tikkanen, M.J. (1994) Low blood total cholesterol and mortality: causality, consequence and confounders. *Am. J. Cardiology*, 73, 80–85

Liu, K., Cooper, R., McKeever, J., McKeever, P., Byington, R., Soltero, I., Stamler, R., Gosch, F., Stevens, E. & Stamler, J. (1979) Assessment of the association between habitual salt intake and high blood pressure: methodological problems. *Am. J. Epidemiol.*, 110, 219–226

Mueller, N., Evans, A., Harris, N.L., Comstock, G.W., Jellum, E., Magnus, K., Orentreich, N., Polk, B.F. & Vogelman, J. (1989) Hodgkin's disease and Epstein–Barr virus. Altered antibody pattern before diagnosis. *N. Engl. J. Med.*, 320, 689–695

Prentice, R.L. (1986) A case–cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73, 1–12

Sherwin, R.W., Wentworth, D.N., Cutler, J.A., Hulley, S.B., Kuller, L.H. & Stamler, J. (1987) Serum cholesterol levels and cancer mortality in, 361, 662 men screened for the Multiple Risk Factor Intervention Trail. *J. Am. Med. Assoc.*, 257, 943–948

Willett, W.C, ed. (1990) *Nutritional Epidemiology*, New York, Oxford University Press

D.J. Hunter
Harvard School of Public Health, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

General issues of study design and analysis in the use of biomarkers in cancer epidemiology

N. Pearce and P. Boffetta

Other contributions to this volume have discussed sources of variation (see Vineis) and measurement error (see White). In this article, we focus on statistical issues involved in the design and analysis of epidemiological studies that use biomarkers. We do not consider statistical issues of laboratory analyses.

In general, epidemiological research involves studying the external, modifiable causes of diseases in populations (McMichael, 1994, 1995) with the intention of developing preventive interventions. In some instances, this activity can be enhanced by using internal biomarkers to obtain better measurements of internal exposure (dose), disease or individual susceptibility. Statistical issues in studies using biomarkers of exposure are not markedly different from those involved in other epidemiological studies based on measures of external exposure, while studies using biomarkers of disease pose specific problems due to the lack of persistence of some of these markers, and the analysis of interaction is of particular interest in studies using markers of susceptibility. As with other epidemiological studies, the statistical analysis of a study involving biomarkers involves in general: (1) relating a particular disease (or health outcome, such as a marker of early effect) to (2) a particular exposure while (3) controlling for systematic error, (4) assessing interactions with other exposures and (5) assessing the possibility of random error. We will consider each of these five aspects of study design and analysis in turn. We will restrict our discussion to full-scale epidemiological studies, i.e. we assume that the use of biomarkers, be they of exposure, effect or susceptibility, aims to contribute to the elucidation of the causal relationships in human populations between diseases and factors such as external exposures, personal habits, genetic traits and interventions. Issues in the design and analysis of transitional studies, in which the main aim is the validation of the markers themselves, are outside the scope of this chapter.

Measuring disease with biomarkers

Epidemiological studies are usually based on a particular population followed over a particular period of time. Miettinen (1985) has termed this study population the 'base population' and its experience over time the 'study base'. The different epidemiological study designs differ only in the manner in which the study base is defined and the manner in which information is drawn from the study base (Checkoway *et al.*, 1989). Thus, epidemiological studies may involve measuring either incidence or prevalence of disease. This distinction is important when a biomarker is being used to measure either the disease under study or early biological effects that are considered to be valid predictors of disease risk (e.g. Rothman *et al.*, 1995). In particular, many studies measuring disease with biomarkers are of cross-sectional design and measure the prevalence of the disease, which is dependent on its incidence and its duration. Thus, in a study looking at markers of cell damage as an effect of exposure to known or suspected carcinogens, the results would depend on factors such as the turnover of the cells in which the marker is measured or the capacity to repair the damage. For example, there is evidence that chromosomal aberrations caused by some carcinogens, such as arsenic and benzene, last for longer periods than aberrations caused by vinyl chloride, and the reason for this difference is not known (Schwartz, 1990).

Incidence studies

Three measures of disease incidence are commonly used in incidence studies. The (person–time) incidence rate (or incidence density; Miettinen,

1985) is a measure of the disease occurrence per unit time. A second measure of disease occurrence is the cumulative incidence (incidence proportion; Miettinen, 1985) or risk, which is the proportion of study subjects who experience the outcome of interest at any time during the follow-up period. A third possible measure of disease occurrence is the incidence odds (Greenland, 1987), which is the ratio of the number of subjects who experience the outcome to the number who do not experience the outcome. As for the cumulative incidence, the incidence odds is dimensionless, but it is necessary to specify the time period over which it is being measured. In incidence studies involving biomarkers of disease, it is therefore important to consider whether a particular biomarker measures incidence or cumulative incidence.

Corresponding to these three measures of disease occurrence, there are three principal ratio measures of effect that can be used in cohort studies. The measure of primary interest is often the rate ratio (incidence density ratio), which is the ratio of the incidence rate in the exposed group to that in the non-exposed group. A second commonly used effect measure is the risk ratio (cumulative incidence ratio), which is the ratio of the cumulative incidence in the exposed group to that in the non-exposed group. When the outcome is rare over the follow-up period, the risk ratio is approximately equal to the rate ratio. A third possible effect measure is the incidence odds ratio, which is the ratio of the incidence odds in the exposed group to that in the non-exposed group. An analogous approach can be used to calculate measures of effect based on the differences rather than the ratios, in particular the rate difference and the risk difference.

In incidence studies involving biomarkers of disease, it is therefore important to consider whether a particular biomarker measures incidence or cumulative incidence. For example, if the 'disease' under study is hepatitis B virus (HBV) infection, then a survey of the prevalence of HBV markers, in a cohort that has been followed over time, will indicate the cumulative incidence of infection in the cohort (with the exception of those who have died from any cause during follow-up or who no longer show evidence of infection). It will not directly indicate the incidence rate of infections; this would require repeated prevalence surveys over time.

Incidence case–control studies

Incidence case–control studies involve studying all of the incident cases of disease generated by the study base and a control group sampled at random from the same study base. The relative risk measure is the incidence odds ratio; the effect measure that this estimates depends on the manner in which controls are selected. Once again, there are three main options (Pearce, 1993).

One option is to select controls from those who do not experience the outcome during the follow-up period, i.e. the survivors (those who did not develop the outcome at any time during the follow-up period). In this instance, a sample of controls chosen by cumulative incidence sampling will estimate the exposure odds of the survivors, and the odds ratio obtained in the case–control study will therefore estimate the incidence odds ratio in the base population. Controls can also be sampled from the entire base population (those at risk at the beginning of follow-up), rather than just from the survivors (those at risk at the end of follow-up). In such case-base sampling, the controls will estimate the exposure odds in the base population of persons at risk at the start of follow-up, and the odds ratio obtained in the case–control study will therefore estimate the risk ratio in the base population. The third approach is to select controls longitudinally throughout the course of the study (Miettinen, 1976); this is sometimes described as 'risk-set sampling' (Robins *et al.*, 1986), 'sampling from the study base' (the person–time experience; Miettinen, 1985) or 'density sampling' (Kleinbaum *et al.*, 1982). In this instance, the controls will estimate the exposure odds in the study base, and the odds ratio obtained in the case–control study will therefore estimate the rate ratio in the study base.

In incidence case–control studies involving biomarkers of disease, it is therefore important to consider whether a particular biomarker measures incidence or cumulative incidence. These issues determine not only which measure of effect is being estimated, but also which method of control selection is appropriate, and the resulting methods of data analysis.

Prevalence studies

The term prevalence denotes the number of cases of disease existing in the population at the time the study was conducted. If we denote the prevalence

of a disease in the study population by P, and if we assume that the incidence rate is constant over time, the population has reached a 'steady state' and there is no migration into or out of the prevalence pool, then it can be shown (Rothman, 1986) that the prevalence odds is equal to the incidence rate (I) multiplied by the average disease duration (D):

$$P/(1 - P) = I \times D$$

Thus, the prevalence odds is directly proportional to the disease incidence, and the prevalence odds ratio is estimated to be:

$$OR = I_1 D_1 / I_0 D_0$$

An increased prevalence odds ratio may thus reflect the influence of factors that increase the disease incidence and/or factors that increase disease duration. The different mechanisms involved in increasing disease incidence or disease duration are likely to involve different time patterns of exposure and disease (see below), which in turn are likely to require different biomarkers for measurement of the etiologically relevant exposures.

Prevalence case–control studies

Just as an incidence case–control study can be used to obtain the same findings as a full cohort study, a prevalence case–control study can be used to obtain the same findings as a full prevalence study in a more efficient manner. In particular, if obtaining exposure information is difficult or costly (e.g. if it involves serum samples), then it may be more efficient to conduct a prevalence case–control study by obtaining exposure information on all of the prevalent cases of the disease under study and a sample of controls selected at random from the non-cases. In this instance, a sample of controls will estimate the exposure odds of the non-cases, and the odds ratio obtained in the prevalence case–control study will therefore estimate the prevalence odds ratio in the base population, which in turn estimates the incidence rate ratio, provided that the average duration of disease is the same in the exposed and non-exposed groups. Once again, an increased prevalence odds ratio may reflect the influence of factors that increase disease incidence and/or factors that increase disease duration, and the different mechanisms involved are likely to require different biomarkers for measurement of the etiologically relevant exposures.

Measuring exposure with biomarkers

Validity of biomarkers of exposure

There are considerable shortcomings in many currently available biomarkers of exposure, including problems of measuring historical exposures; uncertainties as to what a biomarker is measuring; greater susceptibility to confounding in some instances; problems of application to public health policy (Pearce *et al.*, 1995); the disease process affecting the level of the biomarker; and problems of validity of laboratory measurements (Boffetta, 1995). These issues are covered elsewhere in this volume. In this section we concentrate on issues of study design and analysis when measuring exposure with biomarkers, particularly with regard to time-related exposures with a relatively long induction and latency period between exposure and the subsequent occurrence of disease (as is the situation in most cancer epidemiology studies). The issues we discuss are not unique to a particular study design (cohort studies, case–control studies, cross-sectional studies) but rather apply to all studies in which the etiologically relevant time period involves a relatively long induction and latency period, thereby posing problems with the measurement of exposure during this period.

Time-related exposures

Some biomarkers measure factors that are fixed and do not change over time in an individual, e.g. genetic susceptibility genes that may interact with xenobiotic factors in cancer causation (Rothman, 1995). Other biomarkers measure factors that change over time, e.g. micronutrient levels in serum may change from day-to-day (Willett, 1990).

In studies (both prospective and retrospective) of long-term health effects involving time-related exposures, it is important that the time patterns of the study exposure and of the relevant confounders should be taken into account in the analysis (Pearce *et al.*, 1986). In particular, it is important that the principal exposure under study should be analysed in a time-related manner, taking account of the likely induction and latency periods, and the relative etiological importance of exposure intensity, exposure duration and

cumulative exposure. The simplest approach is to analyse the cumulative exposure in a time-related manner, and this may suffice when the aim is merely to consider whether or not there is an effect of exposure. However, once it has been provisionally assumed that an effect exists, attention then shifts to understanding the nature of the effect. In this context, the temporal pattern of exposure and outcome can be considered by examining the effects of exposures in specific time windows while controlling for time-related confounders and for the effects of exposures in other time windows. A more sophisticated approach is direct fitting of a theoretical model of carcinogenesis (Pearce, 1992), which requires assumptions as to the relevance of the times in which the exposure occurred and in which the marker was measured. This may be particularly relevant in studies including the measurement of the exposure, the marker and the disease, which are mainly aimed at elucidating the role of the marker in the exposure–disease relationship (Schatzkin *et al.*, 1993). An example of this type of study is the investigation of the role of human papilloma virus (HPV) in the etiology of cervical cancer (the 'exposure' in this case being factors such as number of sexual partners and age at first intercourse). In this case, the association with HPV infection, when properly measured with PCR-based assays, is of such a magnitude that there is little concern about the relevance of the time periods (Muñoz *et al.*, 1992). However, for cancers from other organs, the association with HPV infection is less clear-cut, and the relevance of the timing of the infection may be one of the unknown factors modifying the exposure–marker–disease relationship.

Thus, biological measures of time-related exposures must be able to measures changes in exposure levels over time. In particular, stored biological samples may not provide valid measurements of long-term patterns of exposure when there are significant variations in exposure over time, unless samples have been taken repeatedly over the course of the study (Armstrong *et al.*, 1992). If it is not possible to take repeated biological samples over time, then it is essential that the samples that are taken relate to the etiologically relevant time period.

Many currently available biomarkers only indicate relatively recent exposures. For example, it is well known that serum levels of micronutrients reflect recent rather than historical dietary intake (Willett, 1990); given the long induction time of most cancers it is usually exposures between 10 and 30 years previously that are etiologically relevant. While this may not be a limitation in cross-sectional studies, provided that the etiologically relevant time period is close to the time of data collection, it is an important limitation in cohort and case–control studies aiming to assess the effects of historical exposures. Some biomarkers are better than others in this respect (particularly biomarkers for exposure to biological agents), but even the best markers of chemical exposures reflect only the last few weeks or months of exposure. On the other hand, with some biomarkers (e.g. serum levels of TCDD; Johnson *et al.*, 1992) it may be possible to estimate historical levels if the exposure period is known, if the half-life is relatively long (and is known) and if it is assumed that no significant exposure has occurred more recently, or if it is reasonable to assume that exposure levels have remained stable over time.

However, historical information on exposure surrogates will often be more valid than current direct measurements of exposure or dose. This situation has long been recognized in occupational epidemiology, where the use of work history records in combination with a job–exposure matrix (based on historical exposure measurements of work areas rather than individuals) is usually more valid than current exposure measurements (whether based on environmental measurements or biomarkers) because of changes in exposure levels over time (Checkoway *et al.*, 1989). Similar problems may occur in the measurement of other carcinogenic exposures. For example, even the best currently available measures of exposure to tobacco smoke, such as plasma or urinary cotinine, appear to have similar validity to questionnaires for the measurement of current exposures; their very short half-life makes them inferior to questionnaires in the estimation of historical exposures (Pearce *et al.*, 1995). On the other hand, some biomarkers would appear to have value in the validation of questionnaires (Forastiere *et al.*, 1993), which can then be used to estimate historical exposures.

Another example in which timing of sample collection is of great importance is in the case of DNA adducts (Wilcosky & Griffith, 1990). Since most adducts are readily repaired, any measure of

exposure based on DNA adducts will depend on the time between the end of exposure and sample collection; this pattern can then be modified by factors such as the activity of repair enzymes, which in turn may also have an independent influence on the outcome, i.e. may be associated with case–control status. DNA adduct formation and repair are particularly problematic, since:

- the extent to which the amount of adducts measured represents the amount of biologically active adducts formed (as discussed above) is usually not known; and
- in the case of measurements taken during exposure (or immediately thereafter, such as at the end of a working day) it is not known how much of the adducts found would persist long enough to be biologically important.

Analysis based on pooled samples

These issues of the timing of sample collection are of particular concern in analyses based on pooled samples. If the samples were not all taken at the same etiologically relevant time period, then the pooled sample will represent the average of samples taken from a variety of time periods. Thus, there may be considerable misclassification with regard to the exposure levels at the etiologically relevant time period.

Systematic error

The major possible types of systematic error (bias) are the same in traditional epidemiology and in studies involving biomarkers (Boffetta, 1995). The various types of bias can be grouped into three major classes: selection bias, information bias, and confounding (Rothman, 1986). This section is not intended to give a comprehensive review of these types of bias; rather, we will concentrate on issues involving data analysis. One solution is to pool small sets of samples stratified on the basis of time since collection; however, this procedure may substantially reduce the advantages of pooling.

Selection bias

Selection bias involves biases arising from the procedures by which the study participants are chosen from the study base. Selection bias can be avoided by including all of the study base (i.e. a cohort study) and obtaining a response rate of 100%. This is often not practicable, but selection bias can also be controlled in the analysis by identifying factors that are related to subject selection and controlling for them as confounders. The statistical issues involved in controlling for selection bias in the analysis are essentially the same as those involved in controlling for sources of confounding (see below).

Information bias

Information bias involves misclassification of the study participants with respect to disease or exposure status. Thus, the concept of information bias refers to those people actually included in the study, whereas selection bias refers to the selection of the study participants from the study base, and confounding generally refers to non-comparability of subgroups within the study base. The various methodological issues of validity, reproducibility and stability of markers are part of the more general problem of information bias.

Non-differential information bias occurs when the likelihood of misclassification of exposure is the same for both cases and non-cases of disease (or when the likelihood of misclassification of disease is the same for exposed and non-exposed persons). Non-differential misclassification of exposure generally biases the relative risk estimate towards the null value of 1.0 (Copeland *et al.*, 1977). Hence, non-differential information bias tends to produce 'false negative' findings and is of particular concern in studies that find no association between exposure and disease.

Differential information bias occurs when the likelihood of misclassification of exposure is different between cases and non-cases (or the likelihood of misclassification of disease is different between exposed and non-exposed persons). This can bias the observed effect estimate in either direction, either towards or away from the null value.

Information bias can drastically affect the validity of a study. As a general principle, it is important to ensure that the misclassification is non-differential, by ensuring that exposure information is collected in an identical manner in cases and non-cases (or that disease information is collected in an identical manner in the exposed and non-exposed groups). In this situation, the bias is in a known direction (towards the null), and although there may be concern that not finding a significant association

(between exposure and disease) may be due to non-differential information bias, at least one can be confident that any positive findings are not due to information bias. Thus, the aim of data collection is not to collect perfect information, but to collect comparable information in a similar manner from the groups being compared, even if this means ignoring more detailed exposure information if this is not available for both groups. However, it is clearly important to collect information that is as detailed and accurate as possible, within the constraints imposed by the need to ensure that information is collected in a similar manner in the groups being compared.

In general, cross-sectional and case–control studies based on biomarkers of exposure are more prone to differential misclassification than studies based on measurement of external exposure and disease, since the biomarkers may be influenced by the disease itself. This problem is less relevant in prospective studies (or nested case–control studies) in which the marker is measured on biological material collected before the onset of disease, provided that cases diagnosed within a short interval after sample collection are excluded. The fact that the relationships between exposure, marker and disease are in most cases obscure limits the interpretation of the findings of biomarker-based studies with respect to the presence or absence of information bias.

Confounding

Confounding occurs when the exposed and non-exposed groups (in the study base) are not comparable due to inherent differences in background disease risk (Greenland & Robins, 1986) caused by exposure to other risk factors. The concept of confounding thus generally refers to the study base, although as noted above, confounding can also be introduced (or removed) by the manner in which study participants are selected from the study base.

If no other biases are present, three conditions are necessary for a factor to be a confounder (Rothman, 1986). First, a confounder is a factor that is predictive of disease (in the absence of the exposure under study); second, a confounder is associated with exposure in the study base; and third, a variable that is intermediate in the causal pathway between exposure and disease is not a confounder. This latter issue is of particular concern in studies using biomarkers, since the identification of potential confounders depends on previous knowledge of the relationship between these and the relevant variables of exposure and outcome, and such knowledge is, for most biomarkers, very poor.

The most straightforward method of controlling confounding in the analysis involves stratifying the data into subgroups according to the levels of the confounder(s) and calculating a summary effect estimate that summarizes the information across strata. However, it is usually not possible to control simultaneously for more than two or three confounders when using stratified analysis. This problem can be mitigated to some extent by the use of mathematical modelling, but this may in turn produce problems of multicollinearity when variables which are highly correlated are entered simultaneously into the model. For example, serum levels of various micronutrients may be strongly correlated, and multicollinearity may occur when they are included in the same model. This will lead to unstable effect estimates with large standard errors, and may in fact lead to the 'wrong' micronutrient showing the strongest association (negative or positive) with disease. This may be one reason why numerous studies have shown that the consumption of green and yellow vegetables is protective against a range of cancers, but the identification of the specific dietary micronutrients involved has remained elusive (Steinmetz & Potter, 1991).

In general, control of confounding requires careful use of *a priori* knowledge, together with assessment of the extent to which the effect estimate changes when the factor is controlled in the analysis. Most epidemiologists prefer to make a decision based on the latter criterion, although it can be misleading, particularly if misclassification is occurring (Greenland & Robins, 1985). The decision to control for a presumed confounder can certainly be made with more confidence if there is supporting prior knowledge that the factor is predictive of disease.

Misclassification of a confounder leads to a loss of ability to control confounding, although control may still be useful provided that misclassification of the confounder was unbiased (Greenland, 1980). Misclassification of exposure is more problematic, since factors that influence misclassifica-

tion may appear to be confounders, but control of these factors may increase the net bias (Greenland & Robins, 1985).

When appropriate information is not available to control confounding directly, it is still desirable to assess its potential strength and direction. For example, it may be possible to obtain information on a surrogate for the confounder of interest. Even though confounder control will be imperfect in this situation, it is still possible to examine whether the main effect estimate changes when the surrogate is controlled in the analysis, and to assess the strength and direction of the change. Alternatively, it may be possible to obtain accurate confounder information for a subgroup of participants (cases and non-cases) in the study and to assess the effects of confounder control in this subgroup.

A related approach involves obtaining confounder information for a sample of the study base (or a sample of the controls in a case–control study). For example, in a study based on questionnaires, biomarkers may be used to validate questionnaire information in a subgroup of study participants.

The potential for confounding is of major concern in all epidemiological studies, including those using biomarkers. The use of biomarkers of exposure does not reduce the need to control for confounding, and in some instances the use of biomarkers may actually introduce confounding into a study. For example, in a study of lung cancer and PAH exposure in a group of factory workers, if the workers are classified according to PAH exposure on the basis of industrial hygiene monitoring, then the percentage of smokers (and the mean number of cigarettes smoked per day) will usually be similar in the groups with low, medium and high levels of occupational exposure to PAH (since these groups are defined purely on environmental levels of PAH exposure in various job categories and departments, and these exposures will usually be unrelated to cigarette smoking). In this situation, cigarette smoking will not be a major confounder. However, if workers are classified according to PAH-DNA levels, these will indicate total exposure to PAHs from all sources, including cigarette smoking. Thus, cigarette smokers will be more likely to be in the 'high PAH exposure' group, and this group will therefore contain a higher proportion of smokers (and a higher mean number of cigarettes smoked per day) than the medium or low PAH exposure group (since some of the total PAH exposure comes from cigarette smoke). The dose–response will then be confounded by cigarette smoking, and the 'high PAH exposure' group may show a higher lung cancer risk which is not due to PAH exposure, but which is actually due to the other carcinogenic constituents of cigarette smoke (Pearce *et al.*, 1995). One solution is to stratify the analysis on cigarette smoking (as measured by questionnaire), but any confounding control is likely to be imperfect. This is even more of a problem if biomarkers are used to measure tobacco smoking because of the problems of measuring the etiologically relevant constituents of tobacco smoke (as distinct from exposure to PAHs in tobacco smoke). Once again, confounding control is likely to be imperfect and, therefore, to yield results that are still confounded and less valid than those obtained by only considering occupational exposures (using a job–exposure matrix). Furthermore, one can only control for known confounders (e.g. tobacco smoke) and cannot control for unknown confounders that may also be subject to the same types of biases described above. Therefore, it is usually preferable to avoid confounding rather than to attempt to control for it *post hoc* (which is why randomized trials are the preferred method when they are feasible). Thus, it is preferable to consider only occupational exposures to PAHs, using a job–exposure matrix, and not to attempt to measure non-occupational exposures to PAH using biomarkers. Finally, it should be stressed that the above issues have been discussed in terms of studies in which exposure is measured prospectively; the problems are even more acute when historical exposures are being assessed.

Random error

Random error will occur in any epidemiological study, just as it occurs in experimental studies. It is often referred to as 'chance', although it can perhaps more reasonably be regarded as 'ignorance' (Checkoway *et al.*, 1989). Even in an experimental study in which participants are randomized into 'exposed' and 'non-exposed' groups, there will be 'random' differences in background risk between the compared groups, but these will diminish in importance (i.e. the random differences will 'even out') as the study size grows. In epidemiological

studies, there is no guarantee that differences in baseline (background) risk will even out between the exposure groups as the study size grows, but it is necessary to make this assumption in order to proceed with the study (Greenland & Robins, 1986). In practice, the study size depends on the number of available participants and the available resources. Within these limitations, it is desirable to make the study as large as possible, taking into account the trade-off between including more participants and gathering more detailed information about a smaller number of participants (Greenland, 1988).

A major problem with the use of biomarkers of exposure and outcome in cancer epidemiology studies (and particularly in cohort studies) is that of small numbers. Even large multicentre cohort studies often struggle to obtain sufficient numbers to assess risks of rare cancers from occupational exposures (Fingerhut *et al.*, 1991; Saracci *et al.*, 1991). The use of biomarkers may be a major problem in this regard, since the resulting expense and complexity may drastically reduce the study size, even in community-based and nested case–control studies, and therefore greatly reduce the statistical power for detecting an association between exposure and disease. As in traditional epidemiology studies, in studies using biomarkers statistical power is a function of the prevalence of exposure and the magnitude of risk; a biomarker with low prevalence and high relative risk can be evaluated in small populations, whereas a biomarker with low prevalence and low relative risk requires a larger population. The optimal balance between precision and validity depends on a number of considerations, including the relative costs of the various exposure measurement techniques (Greenland, 1988). Thus, for an expensive biomarker to be useful it must be substantially better than less expensive and less invasive approaches. However, it has been argued that the necessary study size in some molecular epidemiology studies may be smaller than in traditional epidemiology (Hulka, 1990a,b; Hertzberg & Russek-Cohen, 1993) because of larger differences in biomarker distribution, identification of subgroups at higher risk, and the use of continuous outcome variables (Boffetta, 1995). While this is true in many cases (e.g. the detection of mutations in critical genes as a marker of increased risk of cancer), in other cases the biomarkers may show a very low (or very high) prevalence, thus requiring large samples to detect a difference between groups (Hulka & Margolin, 1992; Rothman *et al.*, 1995). Finally, it should be noted that some biomarkers are of interest in themselves, rather than functioning as surrogates for other exposures; in particular, no alternative methods exist for measuring markers of genetic susceptibility. Nevertheless, the additional information provided by the use of such markers should still be compared with that provided by alternative, larger studies in which the marker is not used.

An additional consideration in study size estimation in studies using biomarkers is the ratio of the number of assays per individual and the number of individuals in the study (Boffetta, 1995), such as the detection of chromosomal aberrations or sister chromatid exchanges. In this case, studies must be based on adequate numbers of individuals and observations per individual (Hirsch *et al.*, 1984; Whorton, 1985). Many biomarkers show marked variation from day to day within the same individual, and the intra-individual variation may be greater than the interindividual variation (Armstrong *et al.*, 1992). It may therefore be necessary to take a large number of measurements to accurately estimate the average exposure level for each individual; otherwise it may be impossible to detect differences between individuals. For example, for 24-hour urinary sodium, the within-person variation may be three times as high as the between-person variation; it has been estimated that the misclassification resulting from taking only one sample per person would result in a true relative risk of 2.0 being reduced to an observed relative risk of 1.2 (Armstrong *et al.*, 1992). Thus, it may be necessary to take 10–15 24-hour urine samples in order to achieve reasonable accuracy in estimating average individual sodium intake levels.

Interaction

Interaction (effect modification) occurs when the estimate of effect of exposure depends on the level of another factor in the study base (Miettinen, 1974). The term statistical interaction denotes a similar phenomenon in the observed data. The former term will generally be used here. Interaction is distinct from confounding (or selection or information bias) in that it does not represent a bias which should be removed or controlled, but rather

a real difference in the effect of exposure in various subgroups that may be of considerable interest. For example, in a cohort study of passive smoking and asthma in children, the effect estimate for passive smoking might be different in different age groups, or in males and females. The clearest example of interaction is when a factor is actually hazardous in one group and protective in another group. More generally, the risk might be elevated in both groups, but the strength of the effect may vary. A typical example of effect modification in studies using biomarkers is the estimate of the risk of disease due to an external agent in subgroups of the population with a different genetic susceptibility marker, such as the polymorphism for an enzyme implicated in the activation or detoxification of the agent (see Landi & Caporasi, this volume). In this situation, effect modification should be interpreted with considerable care, since the presence of statistical interaction may depend on the statistical methods used. In fact, all secondary risk factors modify either the rate ratio or the rate difference, as uniformity over one measure implies non-uniformity over the other. If the assessment of the joint effects of two factors is a fundamental goal of the study, this can be done by calculating stratum-specific effect estimates. It is less clear how to proceed if statistical interaction is occurring, but assessment of joint effects is not an analytical goal. Some authors (e.g. Kleinbaum *et al.*, 1982) argue that it is not appropriate in this situation to calculate an overall estimate of effect summarized across levels of the effect modifier. However, it is common to ignore this stipulation if the difference in effect estimates is not too great (Pearce, 1989). In fact, valid analytical methods (e.g. standardized rate ratios) have been specifically developed for this situation (Rothman, 1986).

Biomarkers may provide better opportunities for assessing interactions between two or more genetic and/or environmental factors. In particular, genetic susceptibility genes should produce a higher disease risk for exposed susceptible groups than for non-susceptible and non-exposed groups (Boffetta, 1995).

However, a major problem of testing for interaction, e.g. in studies involving markers of genetic susceptibility, is that it usually requires a substantial increase in study size. For example, in a case–control study, testing for interaction involves comparing the sizes of the odds ratios (relating exposure and disease) in different strata of the effect modifier, rather than merely testing whether the overall odds ratio is different from the null value of 1.0. The power of the test for interaction therefore depends on the numbers of cases and controls in specific strata (of the effect modifier) rather than the overall numbers of cases and controls. For example, Smith and Day (1984) give an example of a case–control study that would have to be five times larger to detect a difference between odds ratios of 1.0 and 2.0 in the two different strata of an effect modifier than it would have to be to detect an overall odds ratio of 2.0 (ignoring the effect modifier). In general, when considering possible interactions, the size of the study needs to be at least four times larger than when interaction is not considered (Smith & Day, 1984). Thus, in a study involving markers of genetic susceptibility, the gain in statistical power from considering such markers (thereby yielding higher relative risks in some strata) may be offset by the decrease in statistical power from the need to consider interactions. However, if the exposure of interest is independent from the genetic factor under study, case–case comparisons can be used to study interactions with greater statistical power (Piegorsch *et al.*, 1984).

Conclusions

In some instances, the increasing use of biomarkers in epidemiological studies represents a major improvement in the discipline during the last years (Vineis, 1992). In many cases, biomarkers have helped to improve our knowledge of causes and mechanisms of both disease etiology and prevention. In other cases, however, it is unclear whether they represent an improvement on traditional epidemiological methods. Epidemiological studies based on biomarkers are usually more complex than traditional epidemiological studies, because information is available on a larger number of variables whose biological meaning is often poorly known. The methodological considerations involved in classical epidemiological studies on issues such as measurement of disease, measurement of exposure, selection bias, confounding, precision and interaction also apply to biomarker-based studies, and in most cases the methodological problems of this type of study do not require

solutions different from those used in classical studies. In some cases, however, the use of biomarkers may pose specific problems, which have to be addressed within the general framework of modern epidemiological methods.

Acknowledgements

N.P. is funded by a Professorial Research Fellowship and the Wellington Asthma Research Group is supported by a Programme Grant from the Health Research Council of New Zealand.

References

Armstrong, B.K., White, E. & Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology*. New York, Oxford University Press

Boffetta, P. (1995) Sources of bias, effect of confounding in the application of biomarkers to epidemiological studies. *Toxicol. Lett.*, 77, 235–238

Checkoway, H., Pearce, N. & Crawford-Brown, D. (1989) *Research Methods in Occupational Epidemiology*. New York, Oxford University Press

Copeland, K.T., Checkoway, H., McMichael, A.J. & Holbrook, R.H. (1977) Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.*, 105, 488–495

Fingerhut, M.A., Halperin, W.E., Marlow, D.A., Piacitelli, L.A., Honchar, P.A., Sweeney, M.H., Greife, A.L., Dill, P.A., Steenland, K. & Suruda, A.J. (1991) Cancer mortality in workers exposed to, 2,3,7,8-tetrachlorodibenzop-dioxin. *New Engl. J. Med.*, 324, 212–218

Forastiere, F., Agabiti, N., Dell'orco, V., Pistelli, R., Corbo, G.M., Brancato, G., Pacifici, R., Zuccaro, P. & Perucci, C.A. (1993) Questionnaire data as predictors of urinary cotinine levels among nosmoking adolescents. *Arch. Environ. Health*, 48, 230–234

Greenland, S. (1980) The effect of misclassification in the presence of covariates. *Am. J. Epidemiol.*, 112, 564–569

Greenland, S. (1987) Interpretation and choice of effect measures in epidemiologic analyses. *Am. J. Epidemiol.*, 125, 761–768

Greenland, S. (1988) Statistical uncertainty due to misclassification: implications for validation substudies. *J. Clin. Epidemiol.*, 41, 1167–1174

Greenland, S. & Robins, J.M. (1985) Confounding and misclassification. *Am. J. Epidemiol.*, 122, 495–506

Greenland, S. & Robins, J.M. (1986) Identifiability, exchangeability and epidemiological confounding. *Int. J. Epidemiol.*, 15, 4128

Hertzberg, V.S. & Russek-Cohen, E. (1993) Statistical methods in molecular epidemiology. In: Schulte P. & Perera, F.P, eds, *Molecular Epidemiology: Principles and Practices*. New York, Academic Press, pp 199–216

Hirsch, B., McGue, M. & Cervenka, J. (1984) Characterization of the distribution of sister chromatid exchange frequencies: implications for research design. *Hum. Genet.*, 65, 280–286

Hulka, B.S. (1990a) Overview of biological markers. In: Hulka, B.S., Wilcosky, T.C. & Griffith, J.D., eds, *Biological Markers in Epidemiology*, New York, Oxford University Press, pp. 315

Hulka, B.S. (1990b) Methodologic issues in molecular epidemiology. In: Hulka, B.S., Wilcosky, T.C. & Griffith, J.D., eds, *Biological Markers in Epidemiology*, New York, Oxford University Press, pp. 214–226

Hulka, B.S. & Margolin, B.H. (1992) Methodological issues in epidemiologic studies using biological markers. *Am. J. Epidemiol.*, 135, 200–209

Johnson, E.S., Parsons, W., Weinberg, C.R., Shore, D.L., Mathews, J., Patterson, D.G. & Needham, L.L. (1992) Current serum levels of 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin in phenoxy herbicide applicators and characterization of historical levels. *J. Natl. Cancer Inst.*, 84, 1648–1653

Kleinbaum, D.G., Kupper, L.L. & Morgenstern, H. (1982) *Epidemiologic Research. Principles and Quantitative Methods*, Belmont, CA, Lifetime Learning Publications

McMichael, A.J. (1994) Invited commentary – 'Molecular epidemiology': new pathway or new travelling companion? *Am. J. Epidemiol.*, 140, 1–11

McMichael, A.J. (1995) Re: invited commentary—Molecular epidemiology': new pathway or new travelling companion? *Am. J. Epidemiol.*, 142, 225

Miettinen, O.S. (1974) Confounding and effect modification. *Am. J. Epidemiol.*, 100, 350–353

Miettinen, O.S. (1976) Estimability and estimation in case–referent studies. *Am. J. Epidemiol.*, 103, 226–235

Miettinen, O.S. (1985) *Theoretical Epidemiology*, New York, Wiley and Sons

Muñoz, N., Bosch, F.X., Sanjose, S., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C., Ascunce, N., Gonzalez, L.C., Kaldor, J.M., Guerrero, E., Lorinez, A. Santamaria, M., Alsonso de Ruiz, P., Aristizabal, N. & Shah, K. (1992) The causal link between human papillomavirus and invasive cervical cancer: a population-based case–control study in Colombia and Spain. *Int. J. Cancer*, 52, 743–749

Pearce, N. (1992) Methodological problems of time-related variables in occupational cohort studies. *Rev. Epidemiol. Santé Publ.*, 40, S43–S54

Pearce, N. (1993) What does the odds ratio estimate in a case–control study. *Int. J. Epidemiol.*, 22, 1, 1189–1192

Pearce, N.E. (1989) Analytic implications of epidemiological concepts of interaction. *Int. J. Epidemiol.*, 18, 976–980

Pearce, N.E., Checkoway, H. & Shy, C.M. (1986) Time-related factors as potential confounders and effect modifiers in studies based on an occupational cohort. *Scand. J. Work Environ. Health*, 12, 97–107

Pearce, N., Sanjose, S., Boffetta, P., Kogevinas, M., Saracci, R. & Savitz, D. (1995) Limitations of biomarkers of exposure in cancer epidemiology. *Epidemiology*, 6, 190–194

Piegorsch, W.W., Weinberg, C.R. & Taylor, J.A. (1994) Non-hierarchical designs for assessing susceptibility in population-based case–control studies. *Stat. Med.*, 13, 153–162

Robins, J.M., Gail, M.H. & Lubin, J.H. (1986) More on biased selection of controls for case-control analyses of cohort studies. *Biometrics*, 42, 1293–1299

Rothman, K.J. (1986) *Modern Epidemiology*, Boston, Little, Brown

Rothman, N. (1995) Genetic susceptibility biomarkers in studies of occupational and environmental cancer: methodologic issues. *Toxicol. Lett.*, 77, 221–225

Rothman, N., Stewart, W.F., Rabkin, C.S. & Schulte, P.A. (1995) The development, validation, and application of biomarkers for early biologic effects. In: Mendelsohn, M.L., Peeters, J.P. & Normandy, M.J., eds, *Biomarkers and Occupational Health: Progress and Perspectives*, Washington, D.C., Joseph Henry Press, pp. 109–115

Saracci, R. (1984) Assessing exposure of individuals in the identification of disease determinants. In: Berlin, A., Draper, M., Hemminki, K. & Vainio, H., eds, *Monitoring Human Exposure to Carcinogenic and Mutagenic Agents*, (IARC Scientific Publications No. 59), Lyon, International Agency for Research on Cancer

Saracci, R., Kogevinas, M., Bertazzi, P., *et al.* (1991) Cancer mortality in an international cohort of workers exposed to chlorophenoxy herbicides and chlorophenols. *Lancet*, 338, 1027–1032

Schatzkin, A., Freedman, L. & Schiffamn, M. (1993) An epidemiologic perspective on biomarkers. *J. Int. Med.*, 233, 75–79

Schwartz, G.G. (1990) Chromosome aberrations. In: Hulka, B.S., Wilcosky, T.C. & Griffith, J.D., eds, *Biological Markers in Epidemiology*, New York, Oxford University Press. pp. 146–172

Siemiatycki, J., Wacholder, S., Dewar, R., Wald, L., Bégin, D., Richardson, L., Rosenman, K. & Gérin, M. (1988) Smoking and degree of occupational exposure: are internal analyses in cohort studies likely to be confounded by smoking status? *Am. J. Ind. Med.*, 13, 59–69

Smith, P.G. & Day, N.E. (1984) The design of case control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.*, 13, 356–365

Steinmetz, K.A. & Potter, J.D. (1991) Vegetables, fruit and cancer: I. epidemiology. *Cancer Causes Control*, 2, 325–357

Vineis, P. (1992) Uses of biochemical and biological markers in occupational epidemiology. *Rev. Epidem. Sante Publ.*, 40, S63–S69

Whorton, E.B. (1985) Some experimental design and analysis considerations for cytogenetics studies. Environ Mutagen; 4 (Suppl), 9–15

Wilcosky, T.C. & Griffith, J.D. (1990) Applications of biological markers. In: Hulka, B.S., Wilcosky, T.C. & Griffith, J.D., eds, *Biological Markers in Epidemiology*, New York, Oxford University Press, pp. 16–27

Willett, W. (1990) *Nutritional Epidemiology*. New York, Oxford University Press

Corresponding author
N. Pearce
Wellington Asthma Research Group,
Wellington School of Medicine, PO Box 7343,
Wellington, New Zealand

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Sources of variation in biomarkers

P. Vineis

Epidemiology is interested in variation. The goal of epidemiological research is to infer cause–effect relationships by observing whether the occurrence of disease varies according to relevant exposures. Epidemiology is usually interested in 'intergroup' variation (e.g. between those who are exposed and those who are unexposed to the factor of interest), while 'intragroup' variation is a source of noise. Therefore, the study design aims to increase intergroup variability, to limit intragroup variability and to reduce error by ensuring validity and reliability of measurements. The goal of this chapter is to summarize simple ways to recognize the main sources of measurement error and to estimate the extent and the impact of such error.

In fact, the impact of error in the measurement of biomarkers within epidemiological research may be dramatic. For example, it has been claimed that in the measurement of oxidative damage to DNA, routine phenol-based DNA purification procedures can increase 8-hydroxydeoxyguanosine levels 20-fold in samples that are exposed to air following removal of the phenol. Exposure to air alone results in a fourfold increase compared with DNA samples that have been solubilized in buffers purged with nitrogen (Wiseman *et al.*, 1995). It is obvious that such gross contamination would seriously hamper an epidemiological study. The consequences would be even greater if subsets (batches) coming from different subgroups in the study population (e.g. exposed versus unexposed) undergo different technical procedures with different levels of error.

An example of interlaboratory variability, concerning the measurement of DNA adducts in the same samples, is shown in Table 1.

To achieve an accurate estimate of the association between any marker and disease, in epidemiology we need reliable and valid measurements of exposure, covariates (potential confounders and effect modifiers), and outcomes. Causal inference is impossible in the absence of such requirements. I will distinguish, in the following, between marker (any variable that can be measured and is informative for the purposes of the study), assay (a specific laboratory test which aims at measuring that marker) and measurement (the concrete act of measuring the value of a marker in an individual using a specific assay). For example, PAH-DNA adducts are a type of marker, ^{32}P-postlabelling is a type of assay, and the data shown in Table 1 are measurements.

Validity and reliability

Validity is defined as the (relative) lack of systematic measurement error when comparing the actual observation with a standard—a reference method that represents the 'truth'. Such 'truth' is in fact the abstract concept we are interested in, e.g. 'cancer' as defined through an histological characterization. While validity entails a 'standard', reliability concerns the extent to which an experiment or any measuring procedure yields the same results on repeated trials (Carmines & Zeller, 1979). By 'the same results' we do not mean an absolute correspondence, but a relative concept, i.e. a tendency towards consistency found in repeated measurements of the same phenomenon. Reliability is relative to the type and purpose of the measurement: for some purposes we may accept a level of unreli-

Table 1. ^{32}P-postlabelling of aromatic DNA adducts (per 10^8 bases ± SD) from foundry workers

	Laboratory		
	1	2	3
Control (n = 6)	3.1 ± 1.7	–	1.7 ± 0.7
Exposed workers (n = 35)	26 ± 43	12 ± 10	9.2 ± 23

Data from Savela *et al.*, 1989.

ability that is unacceptable for other purposes. In addition to being reliable, the marker must be valid, i.e. it must provide an accurate representation of some abstract concept.

Reliability focuses on a particular property of empirical markers—the extent to which they provide consistent results across repeated measurements. Validity concerns the crucial relationship between concept and marker. Both are a matter of degree, not an 'all or none' property. Validity and reliability are independent: a measurement may be perfectly reliable (reproducible in different laboratories and repeatable at different times) but consistently wrong, i.e. far away from the true value. For example, a gun may be completely reliable if all the shots take place within a small area, but seriously biased if this area is far away from the target; conversely the gun is unbiased but unreliable if the shots have an average distribution around the centre of the target but are dispersed in a large area. We are interested in both validity and reliability; however, since validity is often not measurable, reliability is sometimes used (incorrectly) as a surrogate.

An aspect that is clearly relevant to the discussion of measurement error is timing. Any inference about the meaning of biomarker measures should be strictly time-specific, since time influences the results in several different ways.

The major components of biomarker variability that affect the design of epidemiological studies are variability between subjects (intersubject), variability within subjects (intrasubject) and variability due to measurement errors. The impact of these three categories of variability on the biomarker response can be represented by a linear model of the following form (Taioli *et al.*, 1994):

$$Y_{ijk} = u + a_i + b_j + e_{ijk}$$

where Y_{ijk} is the marker response for subject i at time j and replicate measurement k; u is the true population mean response; a_i is the offset in mean response for subject i (assumed to be normally distributed with mean = 0 and variance = s_i^2; the variance represents the extent of intersubject variability), b_j is the offset in response at time j (assumed to be normally distributed with mean = 0 and variance = s_j^2; this variance represents the extent of intrasubject variability), and e_{ijk} is the assay measurement error (normally distributed with mean = 0 and variance = s^2_{ijk}) (Taioli *et al.*, 1994). The normality of distribution, assumed in the model, must be verified. In fact, many biomarkers have distributions that are far from being normal. Normalization can be achieved through an appropriate transformation, e.g. log transformation.

The model is based on a linear (additive) assumption, which implies that measurement errors are independent of average measurements. Such an assumption must be verified case by case, e.g. by checking whether errors are correlated with the mean.

Impact of measurement error: random and systematic error

Errors of marker measurement may have different impacts depending on the error distribution. If the epidemiological study has been conducted blindly, i.e. laboratory analyses have been done with no knowledge of the exposed/unexposed or diseased/healthy status of the subjects, we expect that the measurement error will be evenly distributed across strata of exposure or disease. Table 2 shows an example. Suppose we recruit a cohort of smoking fathers, in order to investigate respiratory disease among their children. We determine urinary nicotine and cotinine only at recruitment (T_0), and not at other time points during follow-up. However, a number of fathers change their smoking habits in the course of time: at time T_1, 250 smokers have ceased, while 150 non-smokers have started smoking, so that at time T_2, there are 1100 smokers and 850 non-smokers. Therefore, we have 'misclassification' of exposure, based on our spot measurement of urinary nicotine. Misclassification depends, in this case, on the study design and the characteristics of the population, not on technical limitations of the test (assay). The effect of misclassification is such that a 'true' ratio of 2.0 between the prevalence of respiratory symptoms among children born from smokers and the prevalence among children born from non-smokers becomes an 'observed' ratio of 1.48 (Table 2). This underestimation of the ratio is due to a 'blurring' of the relationship between exposure and disease, which occurs when misclassification of exposure, as in this case, is evenly distributed according to the diseased/healthy status (i.e. is not influenced by the outcome).

Both underestimation and overestimation of the association of interest may occur when misclassification is not evenly distributed across the study variables. In the example, we may have a more general distortion of the etiological relationship, and not only a blurring of the association, if classification of exposure depends on the outcome (diseased/healthy status). Blurring is 'bias toward the null', while distortion as a consequence of uneven distribution of misclassification can be in either direction, both towards and away from the null hypothesis.

A model of study design used in large cohort studies is the nested case–control study, which consists of identifying, in the course of the follow-up period, all the subjects who develop the disease and drawing a sample of the healthy subjects (see Pearce & Boffetta, this volume). The advantage of such a design is that with less effort we may have estimates that are as stable as those that would be obtained by analysing the whole cohort. This is particularly useful when measuring biochemical/molecular markers. In our example, suppose we collect urine samples from all 1950 fathers and store them in a refrigerator. When a case of respiratory disease is diagnosed in a child, we take the father's urine sample and, say, the urine samples of two fathers whose children are healthy at that time (cases and controls are usuallymatched for the time-lag between sample collection and sample analysis, to take analite degradation into account; see below). In such a situation, bias due to uneven distribution of error is a potential issue. For example, samples from fathers of diseased children might be analysed in a more accurate way than samples from fathers of healthy children. Scientific curiosity might lead one to measure additional biomarkers among the fathers of diseased children (including, for example, DNA adducts in exfoliated bladder cells or markers of oxidative damage to DNA), on the assumption that these subjects are scientifically more interesting. This is a rather common attitude, which may, however, lead to serious bias. Bias can go both ways. In our example, if oxidative damage to DNA is measured more accurately among case subjects (i.e. if samples are solubilized in buffers purged with nitrogen) than among controls, we might find higher levels in the latter.

A more realistic example of bias dependent on knowledge of the disease status by the researcher is related to degradation of analites when biological samples are stored for a long time. If the samples from the cases affected by the disease of interest and those from controls (within a cohort design) are analysed at different times, bias can arise from differential degradation in the two series. For example, the researcher may decide (incorrectly) to analyse samples from the cases as soon as these arise in the cohort, while controls are analysed at the end of the study. Since the levels of, say, vitamin C decrease rapidly with time, serious bias may arise from the differential timing of measurement in the two series. For this reason, biochemical analyses should be made after matching of cases and controls for time since sample collection.

Table 2. Hypothetical cohort of 1950 fathers (smokers and non-smokers), each with two children.

	T_0	T_1	T_2
Smokers	1200	250 quit	1100
Non-smokers	750	150 start	850

The true prevalence of respiratory symptoms at time T_2 among the children is 10% among children of smokers (220/2200 children) and 5% among children of non-smokers (85/1700 children) (prevalence ratio = 2.0).
Estimated prevalence of respiratory symptoms among those classified as smokers at T_0: 10% (true prevalence in smokers) × 950 + 5% (prevalence in non-smokers) × 250 = 107.5/1200 = 8.9%.
Estimated prevalence of respiratory symptoms among those classified as non-smokers at T_0: 10% (true prevalence in smokers) × 150 + 5% (prevalence in non-smokers) × 600 = 45/750 = 6%.
Estimated prevalence ratio = 6.9%/6% = 1.48.

Sources of variability: intersubject, intrasubject, laboratory

Variation in biomarkers includes interindividual (intersubject) variation, intrasubject variation (i.e. variation in the marker over a particular time period), sampling variation (i.e. variation depending on the framework of biological sample collection) and laboratory variation. Sometimes intra-individual and/or sampling variation are so large that the laboratory measurement variation makes a marginal

Table 3. Correlation coefficients between serum levels of vitamins (first measurement) and subject characteristics

	Weight	Age
Lutein	−0.06 (0.63)	0.08 (0.48)
Criptoxanthine	−0.22 (0.05)	−0.04 (0.69)
Lycopene	−0.25 (0.03)	0.04 (0.73)
α-Carotene	−0.12 (0.29)	−0.03 (0.78)
β-Carotene	−0.27 (0.02)	0.11 (0.32)
Tocopherol	−0.12 (0.30)	0.08 (0.49)

Measurements were made by Giuseppe Malani, Istituto Nazionale della Nutrizione, Roma.
P values are in brackets ($n = 72$).

contribution to overall variation. A particular example of intrasubject variation is associated with error due to handling, processing and storing of specimens; such variability can be measured only if repeated samples from the same individual are collected.

Intersubject variability in marker response may derive from factors such as ethnic group, gender, diet or other characteristics. For example, Table 3 shows interindividual variation concerning the levels of serum vitamins according to weight and age. We might have expected that vitamin levels varied according to diet-related characteristics such as weight (thinner subjects are expected to have healthier diets and higher vitamin consumption; P. Vineis, unpublished data). In this particular example, age is not associated with vitamin levels, although age always has to be considered as an important source of intersubject variation.

Similarly, the marker response may vary within the same subject over time due to changes in diet, health status, variation in exposure to the compound of interest (for dietary items, season is an important variable), and variation in exposure to other compounds that influence the marker response. Table 4 shows intra-individual seasonal variation in levels of micronutrients, indicating that temperature and rainfall (as markers of season) had an important effect on most measurements.

Biological sampling variation is related to circumstances of biological sample collection. For example, hyperproliferation of colonic cells is extremely variable in different points of the colon mucosa. Therefore, not only is intrasubject variation over time important, due to variable exposure to agents that induce cell proliferation, but measurements are also are strongly influenced by how and where the mucosa is sampled. For example, a study (Lyles *et al.*, 1994) estimated that 20% of the variability of the rectal mucosa proliferation index (measured by nuclear antigen immunohistochemistry) is due to subject, 30% is due to the biopsy within the subject, and 50% is due to crypts within a biopsy. In other words, as much as 80% of variation is related to sampling.

Laboratory measurements can have many sources, in particular two general classes of laboratory errors: those that occur between analytical batches and those that occur within batches. An example of a study that was designed in order to assess the different sources of laboratory variation is reported by Taioli *et al.* (1994), using the model described above. In one experiment, they drew blood from five subjects three times in three different weeks ($n = 5$, $k = 3$, $j = 3$) in order to measure DNA–protein cross-links. Table 5 shows the results, which indicate that variation between batches is quite important and greater than variation between subjects. An interaction between intersubject variation and batch variation is suggested.

Methodological issues should be discussed as much as possible within biomarker categories, due to the specificities of each category. Table 6 shows how methodological data can be organized according to biomarker type. Intra-individual and sampling variations are considered, due to the extent of their influence on actual measurements for most markers.

Measurement of variation

Reliability

The extent of variability in measurements can itself be measured in several ways. Let us distinguish between continuous measurements and categorical measurements. A general measure of the extent of variation for continuous measurements is the coefficient of variation (CV = standard deviation/mean, expressed as a percentage). A more useful measure is the ratio between CV_b and CV_w; CV_w measures the extent of laboratory variation within the same sample in the same assay, CV_b measures

the between-subject variation, and the CV_b/CV_w ratio indicates the extent of the between-subject variation relative to the laboratory error. Large degrees of laboratory error can be tolerated if between-person differences in the parameter to be measured are large.

A frequently used measure of reliability for continuous measurements is the intraclass correlation coefficient, i.e. the between-person variance divided by the total (between- plus within-subject) variance. The intraclass coefficient is equal to 1.0 if there is exact agreement between the two measures on each subject (thus differing from the Pearson correlation coefficient which takes the value 1.0 when one measure is a linear combination of the other, not only when the two agree exactly). A coefficient of 1.0 occurs when within-subject variation is null, i.e. laboratory measurements are totally reliable. The intraclass correlation coefficient can then be used to correct measures of association (e.g. relative risks, RRs) in order to allow for laboratory error. Table 7 shows an example where intraclass correlation coefficients were computed for different laboratories which measured estrone levels, and where these coefficients were used to estimate the effect of laboratory error on the observed relative risks (for true relative risks of 1.5, 2.0 and 2.5). In fact, labora-tory error attenuates the causal relationship between the parameter of interest (estrone) and disease (breast cancer), as a consequence of random misclassification. De-attenuation of the relationship (i.e. a correction for measurement error) can be achieved by applying the intraclass correlation coefficient. In Table 7, a high intraclass correlation coefficient (Lab. 4, $r = 0.90$) corresponds to good agreement between true and observed relative risks, while low coefficients (Lab. 1, $r = 0.12$) have a dramatic effect on observed relative risks, which are all around 1.1. The observed RR was obtained using the formula $RR = \exp(\ln RR_t \times r)$, where RR_t is the true relative risk.

Table 4. Correlation over time for some micronutrients, according to environmental variables

	Temperature	Rainfall
Lutein	−0.89[a]	0.74
α-Carotene	0.54	−0.17
β-Carotene	0.56	−0.29
Retinol	−0.69	0.80[b]
α-Tocopherol	−0.60	0.47

From Cooney *et al.*, 1995; changes in seasonal levels were compared for each 2-month period for 21 individuals over a 1-year period; Pearson correlation coefficients.
[a] $P \leq 0.02$.
[b] $P \leq 0.05$.

The intraclass correlation coefficient can be used to estimate the extent of between-subject variability in relation to total variability. The latter includes variation due to different sources (reproducibility, repeatability and sampling variation). To measure reproducibility, i.e. the ability of two laboratories to agree when measuring the same analite in the same sample, the mean difference between observers (and the corresponding confidence interval) has been proposed (Brennan & Silman, 1992). Let us consider the example in Table 8: two pathologists interpeted the slides of 40 subjects with bladder cancer, in order to quantify *p53* overexpression as measured by immunohisto-

Table 5. Analysis of variance in a reliability study: DNA–protein cross-link

Variance component	Variance estimate	*F*-test (*df*)	*P*
Week	0.05	13.7 (2,7)	<0.01
Between subject	0.02	3.4 (4, 7)	>0.05
Week × subject	0.01	2.3 (7, 40)	0.045
Error	0.03		

From: Taioli *et al.*, 1994.
df, degrees of freedom.

Table 6. Organization of methodological issues according to biomarker type

Biomarker category	Intra-individual variation	Biological sampling variation
Internal dose (blood)		
Hormones	Yes (diurnal variation)	No
Water-soluble nutrients	Yes (short half-life)	No
Organochlorine	No (long half-life)	No
Biologically effective dose		
Peripheral white blood cells	Yes (half-life: weeks to months)	No
Exfoliated urothelial cells DNA adducts	Yes (half-life: months)	Yes
Early biological effects		
Lymphocyte metaphase chromosome aberrations	More or less stable	?
Somatic cell mutations glycophorin A	Probably low	No (?)
Intermediate markers		
Cervical dysplasia	Yes	Yes
Colonic hyperproliferation	Yes	Yes
Genetic susceptibility		
Genotype assay	No	No
Non-inducible phenotype	No	No
Inducible phenotype	Yes	No
Tumour markers	Yes	Yes

chemistry. Table 8 gives the percentage of positive-staining cells according to each pathologist and the difference between the two observers for each patient. The proposed measure of interobserver agreement is the mean difference (in this case 0.9) and the corresponding 95% confidence interval, which is between –20.9 and +22.7 (I have used 2 as the probability point of the *t*-distribution, although the number of observations is only 37; in theory this choice would be appropriate only with $n > 50$). What we can conclude is that there is good average agreement between the two pathologists (mean difference = 0.9), but a large confidence interval indicating large overall variability. In the case of immunohistochemistry, a very important source of variation is sampling, since pathologists usually do not read the same fields, i.e. the same cells, at microscope.

Let us now consider categorical (binary) data. Four different pathologists have read specimens of bladder cancer, specially stained to reveal the concentration of the p53 protein (an immunohistochemical technique). Accumulation of the p53 protein (i.e. overexpression of the *p53* tumour suppressor gene) is considered to be a feature of malignancy. Data are expressed as proportions of cells which stain positive, i.e. a continuous variable. Therefore, a simple measure of agreement between readers is the correlation coefficient (Table 8). All coefficients are strongly statistically significant ($P < 0.0001$) and range from 0.6 to 0.9. However, for clinical purposes *p53* expression is generally used in a categorical (binary) way, as in Table 9. Some researchers use 20% as a cut-off point to suggest a more aggressive clinical strategy. If we compare the first and the second pathologist, we see that one counts 11/37 positive cases, and the other 16/37 (three have been considered as not interpretable for technical problems). The discrepancy between the two proportions is more informative than the simple correlation coefficient, since it suggests that at least five patients would receive different clinical suggestions on the basis of the two readings. However, if we look at the detail of the table, we see that the two pathologists agree about the positivity (≥20%) in only 10 cases; in the other 20 they agree on the lack of positivity (<20% of stained cells), while for seven patients there is disagreement.

Simply to state that agreement concerns 30/37 readings is insufficient, because agreement may arise as a consequence of chance. Let us suppose that the two pathologists completely ignore the process underlying their observations, or even throw dice to guess the p53 positivity. Nevertheless, they will agree in a number of cases. Let us imagine that the two pathologists have a propensity to diagnose 11/37 and 16/37 positive cases, but the allocation of subjects is left to chance alone, and the two readings are completely independent. Then we will have the results shown in Table 10, in which a is $26 \times 21/37 = 15$, and so on (multiplication of marginals is an expression of statistical independence of the two readers). The level of agreement due to chance alone is 20/37. Therefore, agreement beyond chance is $(30 - 20)/37 = 10/37$. However, the total potential agreement between the two readers cannot be 100%, but only $(1 - 20/37)\%$; i.e. to be fair we must subtract chance agreement from 100% to have an estimate of total attainable agreement. The measure $(10/37)/(1 - 20/37) = 0.59$ is called the kappa index and measures agreement beyond chance. In this case, the two pathologists agreed in 59% more cases than they would have agreed on the basis of chance alone, i.e. by throwing dice. There are other measures of agreement beyond chance, and kappa has to be used cautiously since there are some methodological pitfalls; for example, the value of kappa strictly depends on the prevalence of the condition which is being studied; with a high underlying prevalence we expect a high level of agreement (Brennan & Silman, 1992).

In addition to reproducibility, i.e. agreement between readers on the same set of observations, with similar techniques we can measure repeatability, i.e. agreement within the same observer at different times (repeat observations).

Validity

Until now we have considered reliability as a property of the assay in the hands of different readers (reproducibility) or at repeat measurements (repeatability). Let us now consider validity of assessment, i.e. correspondence with a standard. It is essential to bear in mind that two readers may show very high levels of agreement, as measured for example by the Pearson correlation coefficient (i.e. $r = 0.9$), even if the first consistently records twice the value recorded by the second observer. Alternatively (e.g. when using the intraclass correlation coefficient, ICC), two readers could show high levels of agreement (e.g. ICC = 0.9) but poor validity if the same errors repeat themselves for both readers.

Table 7. Intraclass correlation coeffcients (r) for the measurement of estrone by different laboratories and resulting observed relative risks given true relative risks of 1.5, 2.0 and 2.5

		True relative risks		
		$RR_t = 1.5$	$RR_t = 2.0$	$RR_t = 2.5$
Laboratory	r	**Observed relative risks**		
Laboratory 1	0.12	1.1	1.1	1.1
Laboratory 2				
Analysis 1	0.82	1.4	1.8	2.1
Analysis 2	0.53	1.2	1.4	1.6
Laboratory 3				
Analysis 1	0.57	1.3	1.5	1.7
Analysis 2	0.65	1.3	1.6	1.8
Laboratory 4	0.90	1.4	1.9	2.3

Observed RR = exp ($\ln RR_t \times r$).
From Hankinson *et al.*, 1994.

Table 8. Measurement of interobserver variation: immunohistochemistry for the expression of the *p53* gene[a]

Specimen number	Pathologist		
	First	Second	Difference
1	16.00	20.00	−4.00
2	83.00	60.00	23.00
3	18.50	28.00	−9.50
4	19.40	19.40	0.00
5	96.40	72.80	23.60
6	4.90	4.50	0.40
7	–	3.20	–
8	21.00	16.20	4.80
9	11.40	5.40	6.00
10	13.00	26.00	−13.00
11	9.40	6.00	3.40
12	4.60	1.70	2.90
13	9.70	1.80	7.90
14	20.80	32.10	−11.30
15	43.50	49.00	−5.50
16	5.60	12.50	−6.90
17	40.10	28.00	12.10
18	11.50	7.00	4.50
19	14.60	13.50	1.10
20	–	27.00	–
21	7.50	2.80	4.70
22	7.40	21.00	−13.60
23	16.30	4.40	11.90
24	21.00	46.50	−25.50
25	2.90	0.90	2.00
26	29.80	34.00	−4.20
27	82.00	69.50	12.50
28	75.00	75.20	−0.20
29	7.00	24.00	−17.00
30	7.20	8.40	−1.20
31	5.70	0.50	5.20
32	15.00	6.10	8.90
33	0.50	1.00	−0.50
34	18.00	7.50	10.50
35	9.80	11.00	−1.20
36	4.20	10.50	−6.30
37	15.20	18.00	−2.80
38	9.30	22.00	−12.70
39	–	–	–
40	52.60	29.30	23.30

Difference parameters

n	Mean	SD	Minimum	Maximum
37[b]	0.9	10.907	−25.5	23.6

95% range for agreement: mean ± 2 × SD (diff.) = 0.9 ± 2 × 10.9 = −20.9 to 22.7

[a]Percentage of positive cells as measured by two different pathologists in the same samples.
[b]Three missing values. Data kindly provided by Renato Coda ($n = 40$).

In the case of immunohistochemistry, we might decide that one pathologist is the standard, but this would not be an ideal choice unless we are organizing a training course in which younger pathologists are instructed to read specimens. According to the definition of validity, we are interested in the correspondence of the measurement with a conceptual entity, i.e. accumulation of the p53 protein as a consequence of gene mutation (in fact, without a mutation the protein has a very short half-life and rapidly disappears from the cells). Tables 11 and 12 show data on the correspondence between immunohistochemistry and *p53* mutations. Sensitivity of immunohistochemistry is estimated as 85%, i.e. false negatives are 15% of all samples containing mutations; specificity is estimated as 71% , i.e. 29% of samples not containing mutations are falsely positive at immunohistochemistry. A combined estimate of sensitivity and specificity is the area under the receiver-operating curve (ROC), i.e. a curve which represents graphically the relationship between sensitivity and specificity. In the example shown in Fig. 1, the area under the ROC curve is 90.3% (Cordon-Carlo *et al.*, 1994). It is usually believed (Fletcher et al., 1988) that sensitivity and specificity indicate properties of a test irrespectively of the frequency of the condition to be detected (however, this is an assumption that requires to be verified). In the example of Tables 11 and 12, the proportion of samples showing a mutation is high (32/73 = 44%); it would be much lower, for example, in patients with benign bladder

Table 9. As Table 8, but the data are grouped according to categories of immunohistochemical positivity

	Pathologist 1				
	0	<20%	≥20%	Total	
Pathologist 2					
0	0	0	0	0	
<20%	0	20	1	21	(3 NI)
≥20%	0	6	10	16	
Total	0	26	11	37	
Pathologist 3					
0	0	0	0		
<20%	1	15	3		(5 NI)
≥20%	2	6	8		
Pathologist 4					
0	0	0	0		
<20%	0	18	0		(7 NI)
≥20%	1	7	7		

	Pathologist 2			
	0	<20%	≥20%	
Pathologist 3				
0	0	2	1	
<20%	0	20	1	(5 NI)
≥20%	0	3	8	
Pathologist 4				
0	0	1	0	
<20%	0	23	2	(7 NI)
≥20%	0	0	7	

	Pathologist 3			
	0	<20%	≥20%	
Pathologist 4				
0	0	1	0	
<20%	1	21	3	(7 NI)
≥20%	0	3	6	

NI, not interpretable.

Table 10. Expected distribution of observations if Pathologist 1 and Pathologist 2 of Table 9 agree by chance alone

	<20%	≥20%	Total
<20%	15 (a)	6 (b)	21
≥20%	11 (c)	5 (d)	16
Total	26	11	37

conditions or in healthy subjects. A measure that is useful in predicting how many subjects (among those testing positive) are really affected by the condition we aim to detect is the positive predictive value. In the example, among 39 patients testing positive at immunohistochemistry, 27 actually have mutations, i.e. immunohistochemistry correctly predicts mutations in 69% of the positive cases. Let use suppose, however, that the prevalence of mutations is not 44% , but 4.4% (32/730). With the same sensitivity and specificity values (85% and 71%, respectively), we would have a positive predictive value of 11.8%, i.e. much lower. The predictive value is a very useful measure, because it indicates how many true positive cases we will obtain within a population of subjects who test positive with the assay we are applying. However, we must bear in mind that the predictive value is strongly influenced by the prevalence of the condition: a very low predictive value may simply indicate that we are studying a population in which very few subjects actually have the condition we want to identify.

Tables 13 and 14 show another set of estimates of validity, based on the clinical outcome. In this case, the aim is to understand how useful p53 immunohistochemistry may be in predicting the risk of metastases. While in the previous example the conceptual entity to be predicted by immunohistochemistry was *p53* gene mutation, here the conceptual entity is the aggressiveness of malignancy, as expressed by the occurrence of lymph node invasion in the follow-up of patients who were tested for p53 at diagnosis. In this case we are obviously interested in the positive predictive value which, for example, is 60% at stage 1. However, we also want to establish whether immunohistochemistry is worth measuring, i.e. whether it adds to what we already know from other clinical examinations. The best predictor of outcome is usually the clinical stage, based on the tumour size at diagnosis. We see that among stage 1 tumours, seven out of 25 (28%) manifest lymph node metastases in the follow-up. The ques-

Table 11. Validity of p53 immunohistochemistry as compared with mutations in the *p53* gene (bladder cancer patients)

	p53 nuclear reactivity (Immunohistochemistry)			
***p53* mutations by SSCP**	**–**	**+**	**++**	**Total**
No mutation	29	7	5	41
All mutations	5	8	19	32
Total	34	15	24	73

Data from Esrig *et al.*, 1993.
Sensitivity of immunohistochemistry (+ and ++) = 27/32 = 85%.
Specificity = 29/41 = 71%.
Positive predictive value = (8+19)/(15+24) = 27/39 = 69%.

tion is: does p53 immunohistochemistry add any information to simple knowledge of the stage at diagnosis? A way to answer this question is to estimate the *a posteriori* probability of metastases, after measurement of p53, and to compare it to the *a priori* probability (i.e. in the absence of p53 measurement), which in this case is 28%. The *a posteriori* probability is computed using the relationship:

a posteriori odds = *a priori* odds × likelihood ratio

Table 12. Theoretical distribution based on the same data as in Table 11 but with a prevalence of mutations of 4.4%

Mutations	**Immunohistochemistry**		
	–	**+/++**	**Total**
Yes	5	27	32
No	496	202	698
Total	501	229	730

Sensitivity = 27/32 = 85%.
Specificity = 496/698 = 71%.
Positive predictive value = 11.7%.

Tables 13 and 14 show that in stage 1, the measurement of p53 allows the probability of metastases to increase from 28% (*a priori*) to 63% (*a posteriori*), while in stages 2 and 3 the contribution of p53 measurement is totally insignificant.

Laboratory drift, study design, quality controls

When we organize and analyse an epidemiological study employing biomarkers, we want to minimize total intragroup variability, in order to identify intergroup differences (e.g. between exposed and unexposed or between diseased and healthy subjects) if they exist. Total intragroup variation is the weighted sum of intersubject, intrasubject, sampling and laboratory variations, with weights that are inversely correlated to the numbers of subjects, measurements per subject and analytical replicates used in the study design, respectively. Obviously, if we do not have detailed information we cannot adjust for intragroup variation. This is the reason why in epidemiological studies employing biomarkers it is important to collect, whenever possible, (1) repeat samples (day-to-day, month-to-month or year-to-year variation may be relevant depending on the marker); (2) potentially relevant information on subject characteristics that may influence intersubject variation; (3) conditions under which samples have been collected and under which laboratory analyses have been conducted (batch, assay, specific procedures).

Concerning item (3), measurement variation may occur as a consequence of many different aspects which are related not only to the choice of the assay but also to:

- collection of the sample (how and when a blood sample was drawn; the type of test tube utilized; the amount of biological material collected; for some measurements, whether the subject was fasting; avoidance of exposure to light if we are interested in vitamin C);
- processing of the sample (e.g. speed of centrifuging to separate different blood components; use of a gradient to separate lymphocytes);
- storage (in a simple refrigerator at –20°C; at –70°C; in liquid nitrogen at –196°C; for how long);
- laboratory analyses (interlaboratory variation; assay; technician performing the assay; batch; accidental contamination of the sample).

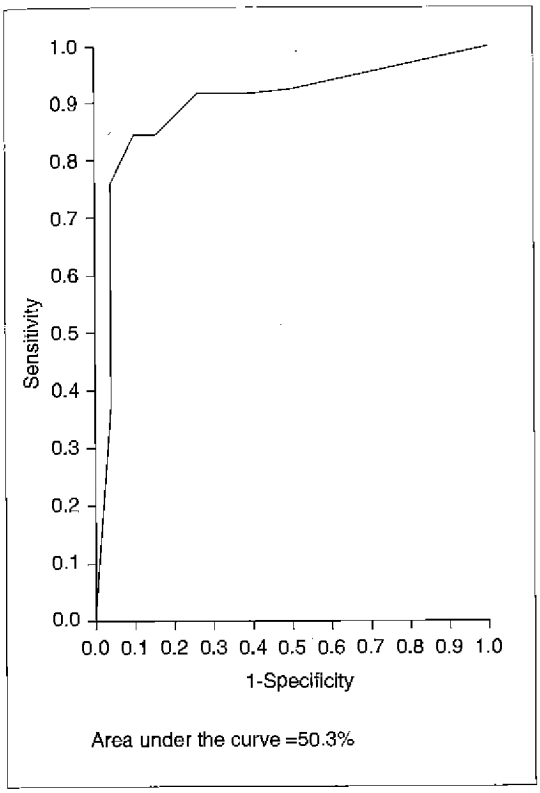

Figure 1. Receiver operating curve (ROC) statistical analysis of the sensitivity and specificity of immunohistochemistry as it relates to PCR–SSCP and sequencing results, representing identification of *p53* mutations. The area under the curve as a measure of diagnostic accuracy was estimated to be 90.3%. Redrawn from Cordon-Carlo *et al.*, 1994.

Table 13. Validity of p53 immunohistochemistry as compared with clinical outcome (occurrence of lymph node metastases during the follow-up)

Stage 1, lymph node invasion	p53 immunohistochemistry		
	Negative	Positive	Total
No	14	4	18
Yes	1	6	7
Total	15	10	25

Data kindly provided by Professor A. Fontana.
Predictive value = 6/10 = 60%.
Likelihood ratio = (6/7)/(4/18) = 3.9
A priori probability = 7/25 = 28%, *a priori* odds = probablity/(1 – probability) = 39%.
A posteriori odds = *a priori* odds × likelihood ratio = 39% × 3.8 = 1.5.

Therefore, in order to minimize intragroup variation, technical details should be considered. As an example, for blood collection the following variables need controlling (Young & Bermes, 1986; Pickard, 1989):

- Collection tubes contamination—e.g. in the case of trace metals, all materials (needles, tubes, pipettes, etc.) should be of a type that does not release metals.
- Types of additive—e.g. collection of plasma entails the use of heparin.
- Order of collection tubes—to avoid carry-over of trace additives, tubes without additives should be proceeded first.
- Time of venipuncture—e.g. measurement of compounds that undergo substantial changes during the day, such as hormones, requires very accurate timing.
- Subject—physiological compounds such as proteins, iron and cholesterol can be increased by 5–15% in the standing position in comparison with supine position.
- Haemolysis may occur as a consequence of tube transport and manipulation.
- Storage conditions.

Laboratory drift is a special problem, but one that is not peculiar to laboratory analyses (e.g. a drift in the quality of interviews typically occurs during longitudinal epidemiological studies). Laboratory drift is a consequence of changes in procedures and accuracy over the course of time, so that the first samples that are analysed tend to differ from subsequent samples. The avoidance of laboratory drift requires a monitoring programme consisting of repeated quality controls. For example, measurements may be compared with a standard at different points in time.

Another source of drift, which cannot be technically avoided, is the degradation of analites when they are stored for a long time. A typical example is given in Table 15, which shows the results of a case–control study nested within a

cohort. After blood collection, as soon as a case of cancer occurred, two suitable controls, matched to the case by time since blood collection, were sampled and their blood was analysed for retinol levels. Table 15 shows that among control subjects the concentration of retinol decreases with the time that has elapsed since blood collection. The impression one gets when looking at retinol levels after less than 1 year since blood collection is that cancer arose in those who had low vitamin levels compared with controls (the difference between cases and controls is statistically significant). However, after 3 years the difference has disappeared, indicating that the first measurement among the cases was likely to have been influenced by the metabolic impairment that occurs in cancer patients. In other words, two events occurred: degradation of retinol in stored samples, which is evident among control subjects; and an inversion of the cause–effect relationship as a consequence of the metabolic impairment in cancer cases. Dealing with this source of drift requires a strict matching of cases and controls for time since blood collection when measurements are planned. In addition, measurements in cases should not be made shortly after blood collection, to avoid the effects of metabolic impairment.

Table 14. *A priori* and *a posteriori* probabilities of lymph node invasion [a]

	A priori probability	*A posteriori* probability (when p53 is positive)
Stage 1	28%	63%
Stage 2	73%	80%
Stage 3	100%	–

[a] Calculated from the same data as in Table 13 (as described in the footnotes of that table) for stages 1, 2 and 3

Conclusions

In the context of epidemiological studies using biomarkers, as well as in 'traditional' questionnaire-based studies, we are interested in both validity and reliability. Validity refers to the truthfulness of estimates, and is therefore the key issue. However, since validity is often not measurable, reliability is sometimes used (incorrectly) as a surrogate. One aspect that is highly relevant to the discussion of measurement error is timing: any inference about the meaning of biomarker measures should be strictly time-specific, since time influences the results in several different ways.

Generally speaking, a blurring of the relationship between marker and disease occurs when errors of measurement are evenly distributed according to the diseased/healthy status (i.e. are not influenced by the outcome). Both underestimation and overestimation of the association of interest may occur when measurement errors are not evenly distributed across the study groups. Blurring is 'bias toward the null', while distortion as a consequence of uneven distribution of measurement errors can be in either direction, both towards and away from the null hypothesis. However, there are some exceptions to this simple rule.

Table 15. Mean retinol levels in the blood of cancer cases and controls

	Time between blood collection and cancer onset					
	Less than 1 year		1–2 years		>3 years	
	n	Mean (µg/l)	*n*	Mean (µg/l)	*n*	Mean (µg/l)
Cases	66	641	45	650	116	694
Controls	132	722[a]	90	701[b]	232	633

From Wald *et al.*, 1986.

[a] $P < 0.001$.

[b] $P < 0.01$.

When we organize and analyse an epidemiological study employing biomarkers, we aim to minimize total intragroup variability and to identify intergroup differences (e.g. between exposed and unexposed or between diseased and healthy subjects). Total intragroup variation is the sum of intersubject, intrasubject, sampling and laboratory variations. Obviously, if we do not have detailed information we cannot adjust for intragroup variation. This is the reason why in epidemiological studies employing biomarkers it is important, whenever possible, to collect repeat samples, potentially relevant information on subject characteristics that influence intersubject variation, and conditions under which samples have been collected and laboratory analyses conducted.

Acknowledgements

I am grateful to Professors Renato Coda, Benedetto Terracini, Enzo Ballatori and Roberto Pagni, and to Drs Stefano Bonassi, Emily White and Nat Rothman for comments and suggestions. This work has been partially funded by the Associazione Italiana per le Ricerche sul Cancro and the Italian National Research Council (Progetto Finalizzato ACRO, Grant No. 93.04716.CT04).

References

Brennan, P. & Silman, A. (1992) Statistical methods for assessing observer variability in clinical measures. *Br. Med. J.*, 304, 1491–1494

Carmines, E.G. & Zeller, R.A. (1979) *Reliability and Validity Assessment*. London, Sage Publications

Cooney, R.V., Frankle, A.A., Hankin, J.H., Custer, L.J., Wilkens, L.R., Harwood, P.J. & Le Marchand, L. (1995) Seasonal variations in plasma micronutrients and antioxidants. *Cancer Epidemiol. Biomarkers Prev.*, 4, 207–215

Cordon-Carlo, C., Dalbagni, G., Saez, G.T., Oliva, M.R., Zhang, Z.F., Rosai, J., Reuter, V.E. & Pellicer, A. (1994) p53 mutations in human bladder cancer: genotypic vs. phenotypic patterns. *Int. J. Cancer*, 56, 347–353

Esrig, D., Spruck C.H., III, & Nichols, P.W. (1993) p53 nuclear protein accumulation correlates with mutations in the p53 gene, tumor grade and stage in bladder cancer. *Am. J. Pathol.*, 143, 1389–1397

Fletcher, R.H., Fletcher, S.W. & Wagner, E.H. (1988) *Clinical Epidemiology—The Essentials*, 2nd edn, Baltimore, Williams and Wilkins

Lyles, C.M., Sandler, R.S., Keku, T.O., Kupper, L.L., Millikan, R.C., Murray, S.C., Bangdiwala, S.I. & Ulshen, M.H. (1994) Reproducibility and variability of the rectal mucosal proliferaton index using proliferating cell nuclear antigen immunohistochemistry. *Cancer Epidemiol. Biomakers Prev.*, 3, 597–605

Hankinson, S.E., Manson, J.E., London, S.J., Willett, W.C. & Speizer, F.E. (1994) Laboratory reproducibility of endogenous hormone levels in postmenopausal women. *Cancer Epidemiol. Biomarkers Prev.*, 3, 51–56

Pickard N.A. (1989) Collection and handling of patients specimens. In: Kaplan, L.A. & Pesce, A.J, eds, *Clinical Chemistry: Theory, Analysis and Correlation*, 2nd edn, CV Mosby, St Louis

Savela, K., Hemminki, K., Phillips, D.H., Hewer, A., Putman, K.L. & Randerath, K. (1989) Interlaboratory comparison of the 32P-postlabelling assay for aromatic DNA adducts in white blood cells of iron foundry workers. *Mut. Res.*, 224, 485–492

Taioli, E., Kinney, P., Zhitkovich, A., Fulton, H., Voitkun, V., Cosma, G., Frenkel, K., Toniolo, P., Garte, S. & Costa, M. (1994) Application of reliability models to studies of biomarker validation. *Environ. Health Perspect.*, 102, 306–309

Wald, N., Boreham, J. & Bailey, A. (1986) Serum retinol and subsequent risk of cancer. *Br. J. Cancer*, 54, 957–961

Wiseman, H., Kaur, H. & Halliwell, B. (1995) DNA damage and cancer: measurement and mechanism. *Cancer Lett.*, 93, 113–120

Young, D.S. & Bermes, E.W. (1986) Specimen collection and processing: sources of biological variation. In: Tiertz, N.W, ed., *Textbook of Clinical Chemistry*, Philadelphia, WB Saunders

P. Vineis
Unit of Cancer Epidemiology, Dipartimento di Scienze Biomediche e Oncologia Umana, Ospedale di S. Giovanni Battista e Università di Torino, via Santena 7, Torino, Italy

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Effects of biomarker measurement error on epidemiological studies

E. White

This chapter presents an overview of how measurement error in a biomarker affects epidemiological studies which use the biomarker. To estimate the effects of biomarker error, one must first measure the error using an appropriate validity or reliability study design and using appropriate parameters, i.e. parameters that are informative about the effects of measurement error on the 'parent' epidemiological study. These measures of the biomarker measurement error from the validity or reliability study can then be applied to what is known about the association under study in the parent study, in order to estimate the effects of the biomarker error on the results of the epidemiological study.

An epidemiological study that uses a biomarker can be of several types. For simplicity, we assume that the biomarker is an intermediate marker between an external exposure and a disease, i.e.:

$$\text{exposure} \rightarrow \text{biomarker} \rightarrow \text{disease}$$

and that the parent epidemiological study assesses the association between the biomarker and the external exposure or between the biomarker and the disease (or between the biomarker and another intermediate marker). Two common study designs in epidemiology that include biomarkers are: (1) a study in which a biomarker is compared between those with a high level of exposure and those with low or no exposure; and (2) a case–control (or nested case–control study) in which the biomarker is compared between those with the disease of interest and controls. Because both types of study are essentially two-group comparisons (either exposure groups or disease groups) on the distribution of the biomarker, these will be discussed together as 'two-group studies'. Measurement error in the biomarker leads to bias in the measure of association (e.g. bias in the odds ratio for the association of biomarker to disease) in the parent study; this bias is called information bias or misclassification bias.

The measurement error for an individual can be defined as the difference between his/her 'measured' (the biomarker 'test') and 'true' biomarkers. The true biomarker can be conceptualized as the underlying biomarker without laboratory or other sources of error, and if the measure can fluctuate over time, the true biomarker would be integrated over the time period of etiological interest (e.g. the mean value over a 10-year period).

There are numerous sources of measurement error in biomarkers. Examples are given in Box 1. Some have been discussed in the previous chapter (Vineis, this volume). Measurement error can be introduced by errors in the laboratory method selected to measure the exposure of interest; for example, if the true biomarker of etiological importance is β-carotene, the choice of total carotenoids in the blood as the biomarker test will include other exposures not relevant to the epidemiological true exposure. Further sources of error include failure to specify fully the protocol in terms of timing and method of specimen collection, specimen handling and storage, and laboratory technique. An additional source of error is the 'random' variation that occurs between batches and between laboratory technicians even when the protocol is well specified.

Of particular importance is the fact that measurement error can be due to short-term (e.g. week to week) biological variation and long-term change in the biomarker within subjects. This type of error is often ignored when assessing laboratory measurement error, but it can have a large impact on an epidemiological study. This is because, unless the biomarker is a fixed characteristic within individuals, the underlying true biomarker

that influences the disease of interest is rarely an individual's measured biomarker on a single day, but rather the average over time. Thus, even a perfect measure of the biomarker at a single point in time could be a poor measure of the 'true' biomarker.

Differential measurement error occurs when the biomarker error differs between those with and those without the disease to be studied in the parent study. A primary concern in case–control studies of biomarker–disease associations is that the biological effects of the disease or treatment may affect the biomarker so that it is no longer an equally good measure of the true (pre-disease) biomarker for both cases and controls. The term differential measurement error can be more generally defined as measurement error that differs between the comparison groups (e.g. disease groups or exposure groups) to be used in the parent study.

Non-differential measurement error occurs when the measurement error does not differ between the comparison groups in the parent epidemiological study. The effects of both differential and non-differential error will be discussed in this chapter.

'Validity' is the relationship between the biomarker test (the mismeasured biomarker) and the true biomarker in a population of interest. Measures of validity are parameters that describe the measurement error in the population. A validity study is defined here as one in which a sample of individuals is measured twice—once using the biomarker test of interest and once using a perfect measure of the true biomarker. To design a validity study that reflects the amount of measurement error that will occur in the parent epidemiological study, several design issues must be considered. First, the sample of individuals for the validity study should be a random sample from the parent study. If this is not possible, the subjects in the validity study should be comparable to the subjects in the parent study in terms of age, sex and other parameters that could influence the distribution (variance) of the biomarker and/or the error of the biomarker test. Second, the biomarker test should be collected, processed, stored and analysed in the validity study using the same procedures that will be used in the parent study. Third, a perfect measure of the true biomarker is needed for all individuals in the validity study. This true measure must reflect the underlying true biomarker without error, including error due to variation in laboratory procedures and variations over time. The last issue is particularly problematic, because the true biomarker of interest is often the average value over many years. In addition, one should consider conducting separate validity studies on those with the disease and those without the disease, to assess whether the measurement error differs between cases and controls. Finally, a validity study should

Box 1. Examples of sources of measurement error in laboratory measures in epidemiological studies

Errors in the laboratory method as a measure of the exposure of interest

Method may not measure all sources of the biological true (etiological) exposure of interest

Method may measure other exposures that are not the true exposure of interest

Method may be influenced by subject characteristics (other than the true exposure) that the researcher cannot manipulate, e.g. by the disease under study or by other diseases

Errors or omissions in the protocol

Failure to specify the protocol in sufficient detail regarding timing and method of specimen collection, specimen handling, storage and laboratory analytical procedures

Failure to include standardization of the instrument periodically throughout the data collection

Errors due to biological variability within subjects

Short-term variability (hour-to-hour, day-to-day) in biological characteristics due to, for example, diurnal variation, time since last meal, posture (sitting versus lying down)

Medium-term variability (month-to-month) due to, for example, seasonal changes in diet

Long-term change (year-to-year) due to, for example, deliberate dietary changes over time

Errors due to variations in execution of the protocol

Variations in method of specimen collection

Variations in specimen handling or preparation

Variations in length of specimen storage

Variations in specimen analysis between batches (different batches of chemicals, different calibration of instrument)

Variation in technique between laboratory technicians

Random error within batch

be analysed using parameters that provide information about the effect of biomarker measurement error on the parent epidemiological study.

Often, a perfect measure of the biomarker does not exist or is not feasible to use in a validity study. Reliability studies, in which repeated measurements of the biomarker are taken on a group of subjects, usually only measure part of the measurement error. However, reliability studies can sometimes be used to measure the validity of a biomarker, without having a perfect measure of that biomarker.

This chapter is divided into six sections. The first looks at quantifying the measurement error in binary biomarkers using a validity study. The second section is a discussion of the effects of the error in a binary biomarker on the odds ratio or other measures of association in the parent epidemiological study. The third considers quantifying measurement error in continuous biomarkers using a validity study, and the fourth section covers the effects of the error in a continuous biomarker on the parent epidemiological study which uses the biomarker. Although one generally needs to conduct a validity study for binary biomarkers, in the case of continuous biomarkers, reliability studies can be designed that yield information about the validity of the biomarker. Thus, information from a reliability study can answer questions about the effect of biomarker error on the parent study. This is covered in the fifth section.

The last section presents techniques to adjust the results of the parent study for the effects of measurement error. The researcher can then estimate the value of the true odds ratio (or other measure of association) that might have been observed in the absence of biomarker error.

Many related topics are beyond the scope of this chapter. A discussion of the effect of measurement error in a susceptibility biomarker (a factor that modifies the relationship between exposure and disease) on the results of the parent study using the susceptibility marker will not be discussed. The reader is referred to other sources for information on the effects of measurement error in a categorical measure (Walker and Blettner, 1985; de Klerk *et al.*, 1989; Armstrong *et al.*, 1992), the effects of measurement error on sample size and power (Armstrong *et al.*, 1992; McKeown-Eyssen, 1994; White *et al.*, 1994), and the design and anaylsis of other types of reliability studies (Dunn, 1989; Armstrong *et al.*, 1992). Techniques to reduce biomarker measurement error, and therefore to reduce the bias in the results of the parent study caused by measurement error, are covered throughout this book.

For the purposes of this chapter, sources of bias other than misclassification bias in epidemiological studies (i.e. bias due to the effects of confounding factors, unrepresentative selection of subjects, and sampling a finite number of subjects) are assumed to be absent. Furthermore, to make possible an explicit analysis of the effects of measurement error for continuous biomarkers, certain simplifying assumptions are made about the form of the error. The purpose is to give the reader some insight into evaluating the effects of measurement errors.

Measuring the error in a binary biomarker

Binary biomarkers are those that classify an analyte or characteristic as present (positive) or absent (negative) for each study subject. Measurement error in a binary (dichotomous) biomarker is usually referred to as misclassification. Binary biomarkers are subject to all of the sources of measurement error as described in the introduction and Box 1.

Measures of misclassification—sensitivity and specificity

The degree of misclassification in a binary biomarker is measured by its sensitivity and specificity. Sensitivity and specificity can be measured in a validity study in which the biomarker under evaluation (the mismeasured biomarker) is compared with a perfect measure of the underlying biomarker (true biomarker) in the population of interest. Individuals are then cross-classified by their results on each test:

		True biomarker	
		+	−
Classified by biomarker test	+	*a*	*b*
	−	*c*	*d*

The sensitivity (sens) of the biomarker under evaluation is the proportion of those who are true positives (positive on the criterion test) who are correctly classified as positive by the biomarker test:

$$\text{sens} = a/(a + c)$$

(Note that the definition given here of sensitivity is different from the meaning in some laboratory contexts, i.e. the lowest level detectable by a test.) The specificity (spec) is the proportion of those who are true negatives who are classified as negative by the biomarker test:

$$\text{spec} = d/(b + d)$$

Even though both sensitivity and specificity can range from 0 to 1, it is assumed that sensitivity plus specificity is greater than or equal to 1. In other words, for the biomarker test to be considered a measure of the true biomarker, the probability that the biomarker test classifies a truly positive person as positive (sensitivity) should be greater than the probability that it classifies a truly negative individual as positive (1 – specificity), i.e. sens > (1 – spec), or (sens + spec) > 1. Thus the parameter (sens + spec – 1), called the Youden misclassification index (Kotz & Johnson, 1988), is a good measure of the total degree of misclassification. If the Youden index is 1, the biomarker test is perfect; if it is zero, the test has no association with the true biomarker; and if it is less than zero, the test is inversely related to the biomarker.

For a validity study to measure the sensitivity and specificity of a biomarker, the study sample can be made up of subjects not restricted by their biomarker status, or one can sample subjects who are true positives and those who are true negatives by the criterion test. However, one cannot sample subjects based on the results of the mismeasured biomarker test and correctly compute sensitivity and specificity.

As noted in the introduction, separate validity studies should be conducted on the comparison groups to be used in the parent study, particularly if one suspects that the sensitivity and/or specificity may differ between groups. Specifically, if the parent epidemiological study is a case–control study in which the disease could influence the biomarker test, then separate validity studies of the biomarker should be conducted on a group with the disease and on a control group. For example, suppose the parent study were a study to measure the presence of a specific protein expression (present versus absent) in normal breast cells among women with concurrent breast cancer compared with those without cancer at the time of a breast biopsy. The true biomarker of interest would be presence of the protein expression in the normal tissue 5–10 years before diagnosis, while the measured biomarker is expression in normal tissue at the time of diagnosis, which could be influenced by the presence of disease. Assuming the protein expression could be measured in a stored tissue sample, the validity study could be conducted on two groups: the subset of cases and the subset of controls who had prior breast biopsies, say, 5–10 years before the index biopsy, with the true measure being expression from the earlier biopsy. This would yield separate measures of sensitivity and specificity for cases and controls.

Effects of the error in a binary biomarker on epidemiological studies

The effect of differential misclassification of a binary biomarker in a two-group study

The effects of misclassification are straightforward for studies in which a binary variable is compared between two groups (Bross, 1954; Newell, 1962; Gullen *et al.*, 1968; Goldberg, 1975; Barron, 1977; Copeland *et al.*, 1977; Fleiss, 1981; Kleinbaum *et al.*, 1982).

In an unmatched two-group study of a binary biomarker, under the assumption that the group status is correctly classified, the effect of misclassification of the biomarker is to rearrange individuals in the true 2 × 2 table into an observable 2 × 2 table. Individuals in the first group (diseased or exposed group) remain in the first group but may be misclassified as to biomarker status, and the second group (non-diseased or unexposed) is also rearranged, as follows:

		True classification	
		Disease or exposure group	
		Group 1	Group 2
		+	–
True biomarker	+	P_1	P_2
	–	$1 - P_1$	$1 - P_2$

Measures of association: Difference $= P_1 - P_2$

$$OR_T = \frac{P_1 (1 - P_2)}{P_2 (1 - P_1)}$$

		Observable classification (misclassification)	
		Disease or exposure group	
		Group 1	Group 2
		+	−
Biomarker test	+	p_1	p_2
	−	$1-p_1$	$1-p_2$

Measures of association Difference = $p_1 - p_2$

$$OR_O = \frac{p_1(1-p_2)}{p_2(1-p_1)}$$

Note that P_1 and P_2 are the true proportions of biomarker-positives in groups 1 and 2 respectively, and similarly p_1 and p_2 refer to the proportions that would be 'observable' as positive by the biomarker test in the two groups. The term observable is used to mean that which would be expected on average when there is measurement error (the actual observed parameter in the parent study would be an estimate of the 'observable' parameter).

There is differential misclassification when the sensitivity of the biomarker test for group 1 ($sens_1$) differs from that for group 2 ($sens_2$) and/or the specificity of the biomarker test for group 1 ($spec_1$) differs from that for group 2 ($spec_2$). The misclassification leads to the observable p_1 and p_2 being different from the true P_1 and P_2 (Goldberg, 1975):

$$\begin{aligned} p_1 &= sens_1 \times P_1 + (1 - spec_1) \times (1 - P_1) \\ p_2 &= sens_2 \times P_2 + (1 - spec_2) \times (1 - P_2) \end{aligned} \quad (1)$$

The first equation states that a proportion ($sens_1$) of the true biomarker-positives (P_1) in the first group plus a proportion ($1 - spec_1$) of the true biomarker-negatives ($1 - P_1$) in the first group will be classified by the biomarker test as positive (p_1) in the first group. The second equation expresses the same concept for group 2.

The association between group and biomarker in the parent study would typically be measured by the difference in the proportion biomarker positive if the two groups were exposure groups and by the odds ratio if the two groups were disease groups. When there is measurement error, these measures of association are biased because they are based on p_1 and p_2:

$$\text{Observable difference} = p_1 - p_2 \quad (2)$$

or

$$OR_O = \frac{p_1(1-p_2)}{p_2(1-p_1)} \quad (3)$$

where p_1 and p_2 are from equations (1).

Differential misclassification can have any effect on the difference: the observable difference, $p_1 - p_2$, could be closer to the null hypothesis or cross over the null hypothesis (i.e. have a different sign) compared with the true difference, $P_1 - P_2$. Similarly, compared with the true odds ratio, the observable odds ratio can be closer to the null hypothesis of OR = 1, be further from the null or cross over the null.

Returning to an earlier example, let us suppose it is hypothesized that a certain protein expression in normal breast tissue is associated with increased risk of breast cancer. This could be tested in a case–control study, using normal tissue removed at the time of diagnosis from breast cancer cases and tissue from controls undergoing a breast biopsy. Suppose that the test was insensitive among both cases and controls, such that only half of those with true protein expression were classified as positive ($sens_1 = sens_2 = 0.5$). Let us assume further that the specificity was perfect for controls ($spec_2 = 1.0$), but that, among cases, 10% of those without expression in normal tissue were classified as positive due to the influence of the breast tumour ($spec_1 = .90$). If there were no true association between breast cancer and the protein expression in normal tissue (e.g. suppose 10% of each group were protein-positive), then applying equations (1) and (3) would yield the following:

True classification:

		Breast cancer	
		+	−
True	+	0.10	0.10
biomarker	−	0.90	0.90
		$OR_T = 1.0$	

Observable classification:

$$p_1 = 0.5 \times 0.1 + 0.1 \times 0.9 = 0.14$$
$$p_2 = 0.5 \times 0.1 + 0 = 0.05$$
$$OR_O = \frac{0.14\ (0.95)}{0.05\ (0.86)} = 3.1$$

Thus, a true odds ratio of 1.0, i.e. no association between the disease and biomarker, could appear as a strong association ($OR_O = 3.1$).

The effect of non-differential misclassification of a binary biomarker in a two-group study

Non-differential misclassification in a two-group study would occur when the sensitivity of the biomarker test is the same for both groups ($sens_1 = sens_2$) and the specificity is the same for both groups ($spec_1 = spec_2$). Then the effect of measurement error in the biomarker on the difference between groups can be computed as in equations (1) and (2), which simplifies to:

$$p_1 - p_2 = (P_1 - P_2)\,(\text{sens} + \text{spec} - 1) \qquad (4)$$

This states that the observable difference between the proportions that are positive in group 1 and group 2, $p_1 - p_2$, based on the biomarker test is equal to the true difference between groups, $P_1 - P_2$, multiplied by a factor equal to sensitivity plus the specificity minus 1. If the biomarker test at a mimimum classifies a true positive person as positive on the test with greater or equal probability than it classifies a true negative person as positive (i.e. the Youden index (sens + spec – 1) is in the range 0–1), then the observable difference is always less than the true difference and does not change sign. For example, if the biomarker test had sens = 0.7 and spec = 0.8, which could be considered reasonable test accuracy, then (by equation 4) the observable difference between groups would only be half of the true difference (i.e. sens + spec –1 = 0.5).

The effect of non-differential misclassification of a biomarker on the odds ratio in a case–control study can be computed by calculating p_1 and p_2 as in equation (1) (except there would be a common sensitivity and a common specificity for the diseased and non-diseased groups) and then calculating the odds ratio (equation 3). As an example, data are adapted from a report by Schiffman & Schatzkin (1994). They report on two case–control studies of human papillomavirus (HPV) infection and cervical intra-epithelial neoplasia (CIN). A major difference between the two studies was the accuracy of the test for HPV. In the first study, a 3 ml cervicovaginal lavage was tested by Southern blot DNA hybridization techniques. In the second study, a 10 ml cervicovaginal lavage was used to collect more adequate DNA specimens, and the specimens were tested by an L1 consensus primer polymerase chain reaction (PCR) technique. The second study is considered here as an epidemiological study without biomarker measurement error, and the first study as one with biomarker error. Had a perfect measure been available at the time of the earlier study, a validity study could have been conducted so that the sensitivity and specificity of the first test could be computed. Adapting data from this report (modified to serve as an example), such a validity study might have yielded the following:

		True (2nd) test	
		HPV+	HPV–
Misclassified (1st)	HPV +	10	6
test	HPV –	5	31
	Total	**15**	**37**

and sensitivity and specificity could be computed as follows:

$$\text{sens} = 10/15 = 0.67$$
$$\text{spec} = 31/37 = 0.84$$

Here we assume that the measurement error is non-differential, i.e. that the sensitivity and specificity are the same for cases and controls. The effect of this magnitude of biomarker error on the true relationship between HPV and CIN can be estimated by applying equations (1) and (3) to the true relationship between HPV and CIN as seen in the second epidemiological study:

		CIN	
		Cases *n* (%)	Controls *n* (%)
True (2nd)	HPV +	381 (0.81)	80 (0.18)
test	HPV –	89 (0.19)	375 (0.82)
		470 (1.0)	455 (1.0)

$$OR_T = \frac{0.81\,(0.82)}{0.18\,(0.19)} = 20$$

$$p_1 = 0.67 \times 0.81 + 0.16 \times 0.19 = 0.57$$
$$p_2 = 0.67 \times 0.18 + 0.16 \times 0.82 = 0.25$$

$$OR_O = \frac{0.57\,(0.75)}{0.25\,(0.43)} = 4.0$$

This shows that a study using the misclassified HPV test would find 57% of cases positive, 25% of controls positive (rather than 81% and 18% based on the accurate test) and an odds ratio of 4.0 rather than the true odds ratio of 20. In fact, the first study using the inaccurate HPV test observed an odds ratio of 3.7.

Non-differential misclassification leads to an attenuation of the odds ratio towards the null value of 1 (Gullen *et al.*, 1968). The degree of attenuation in the observable odds ratio depends not only on the true odds ratio and the sensitivity and specificity of the biomarker test, but also on the proportion of the non-diseased group who are true biomarker-positives (P_2). The observable odds ratio does not 'cross over' the null of 1.

Table 1 gives further examples of the effect of non-differential misclassification on the odds ratio for reasonable values of sensitivity (0.5–0.9), specificity (0.8–0.99) and P_2 (0.1, 0.5) and for true odds ratios of 2 and 4. As can be seen from the table, the attenuation in the odds ratio can be considerable. When the proportion who are truly positive is low (e.g. $P_2 = 0.1$ in upper half of the table), the attenuation of the odds ratio is severe, except when the specificity is very high (e.g. spec = 0.99). When the proportion who are truly positive is high (e.g. $P_2 = 0.5$ in lower half of the table), the observed OR is strongly attenuated from the true OR except when the sensitivity is very high (e.g. sens = 0.9). Even strong associations between the true biomarker and disease would be obscured by moderate values of sensitivity and specificity. For example, for sens = 0.7, spec = 0.8 and $OR_T = 4.0$, the observable odds ratio would be 1.64 for $P_2 = 0.01$ and 1.83 for $P_2 = 0.5$. These observable odds ratios would not be detectable as different from the null value of 1 unless the epidemiological study sample size were large.

Measuring the error in a continuous biomarker using a validity study

Often a biomarker assay yields quantitative information about the amount of an analyte in a biological specimen; these measures can usually be considered to be continuous variables. This section covers the parameters of measurement error that can be derived from validity studies in which each subject in the validity study is measured twice—once using the mismeasured biomarker and once using a perfect (true) measure.

The theory of measurement error in continuous variables and its effects on studies of a continuous outcome were developed in the fields of psychometrics, survey research and statistics (Hansen *et al.*, 1961; Cochran, 1968; Lord & Novick, 1968; Nunnally, 1978; Allen & Yen, 1979; Bohrnsted, 1983; Fuller, 1987). The effects of measurement error have also been derived in the context of epidemiological studies of a continuous exposure variable and a dichotomous disease outcome (Prentice, 1982; Whittemore and Grosser, 1986; Armstrong *et al.*, 1989).

A model of measurement error

A simple model of measurement error in a continuous measure X is:

$$X_i = T_i + b + E_i$$

where $\mu_E = 0$ and $\rho_{TE} = 0$. In this model, for a given individual i, the measured biomarker X_i differs from its true value T_i by two types of measurement error. The first is the systematic error or bias, b, that would occur (on average) for all measured subjects. The second, E_i, is the additional error in X_i for subject i. E will be referred to as the subject error to indicate that it varies from subject to subject. It does not refer just to error due to subject characteristics; rather it includes all of the sources of error outlined in Box 1.

For the population of potential study subjects, X, T and E are variables with distributions, e.g. the distribution of E is the distribution of subject measurement errors in the population of interest. X, T and E would have expectations (population means over an infinite population) denoted by μ_X, μ_T and μ_E, respectively, and variances denoted by σ_X^2, σ_T^2, and σ_E^2. Because the average measurement error in X in the population is expressed as a constant, b, it follows that μ_E, the population mean of the subject error, is zero. The assumption of the model that the correlation coefficient of T with E, ρ_{TE}, is 0 states that the true value of the biomarker is not correlated with the measurement error. In other words, individuals with high true values are assumed not to have systematically higher (or lower) errors than individuals with lower true values.

Table 1. Effect of non-differential misclassification of a binary biomarker on the observable odds ratio (OR_O)

		True OR = 2.0	True OR = 4.0
Biomarker test sensitivity	**Biomarker test specificity**	**OR_O[b]**	**OR_O[b]**
P_2 = 0.1[a]			
0.5	0.80	1.14	1.38
0.7	0.80	1.23	1.64
0.9	0.80	1.32	1.92
0.5	0.90	1.28	1.76
0.7	0.90	1.39	2.09
0.9	0.90	1.48	2.41
0.5	0.99	1.75	3.06
0.7	0.99	1.83	3.33
0.9	0.99	1.89	3.61
P_2 = 0.5[a]			
0.5	0.80	1.24	1.46
0.7	0.80	1.40	1.83
0.9	0.80	1.64	2.59
0.5	0.90	1.35	1.69
0.7	0.90	1.50	2.07
0.9	0.90	1.73	2.85
0.5	0.99	1.48	1.96
0.7	0.99	1.61	2.33
0.9	0.99	1.82	3.11

[a]P_2 is the prevalence of true biomarker-positives in the non-diseased group. P_1, the prevalence of true biomarker-positives in the diseased group, is, by definition, $P_1 = P_2 \times OR_T/[1 + P_2 (OR_T - 1)]$.
[b]OR_O from equations (1) and (3).

Measures of measurement error—bias and validity coefficient

Two measures of measurement error are used to describe the relationship between X and T in the population of interest, based on the above model and assumptions. One is the bias, i.e. the average measurement error in the population:

$$b = \mu_X - \mu_T$$

The bias in X can be estimated from a validity study as:

$$\hat{b} = \bar{X} - \bar{T}$$

The other is a measure of the 'precision' of X, i.e. the variation of the measurement error in the population. One measure of precision is σ_E^2, the variance of E, which is often called the within-subject variance. (Note that the model is formulated so that the two measures of measurement error have separate parameters. The average error is given by b, and b does not contribute to the variance of the error. The error that varies from subject to subject is parametrized by E and measured by σ_E^2, and E does not contribute to the average error.) A more important measure of precision is the correlation of T with X, ρ_{TX}, termed here as the validity coefficient of X. The measure ρ_{TX} is important because it relates the within-subject variance σ_E^2 to the total variance σ_X^2, and it is this ratio, along with the bias, that measures the impact of biomarker error on the parent epidemiological study. Using the above model, it can be shown that ρ_{TX} is the square root of 1 minus the variance of E relative to the variance of X (Allen & Yen, 1979), i.e.:

$$\rho_{TX} = \sqrt{1 - (\sigma_E^2/\sigma_X^2)} \quad (5)$$

Using the above equation, it can be seen that the smaller the error variance, the greater is ρ_{TX}. ρ_{TX} would range between 0 and 1, with a value of 1 indicating that X is a perfectly precise measure of T. ρ_{TX} is assumed to be zero or greater, i.e. for X to be considered to be a measure of T, X must be at a minimum positively correlated to T.

ρ_{TX} can be estimated in a validity study by the Pearson correlation coefficient of X with T. Thus, ρ_{TX}^2 can be interpreted as the proportion of the variance of X explained by T. Using the above model:

$$\rho_{TX}^2 = \sigma_T^2/\sigma_X^2$$

For example, if ρ_{TX} were 0.8, this would reflect that only 64% of the variance in X is explained by T, with the remainder of the variance being due to the error.

To further understand the concepts of bias and precision, consider a situation in which X only has a systematic bias, with $E_i = 0$ for all subjects. For example, suppose that the only source of error in a measure of serum cholesterol was that it quantitated each individual exactly 100 mg/dl too high. Then, in a population, the variable X, even though it has systematic measurement error, could be used to order each person in the population perfectly by their value of T. X would be biased but have perfect precision. However, if E_i varied from person to person (around the mean $\mu_E = 0$), the ordering is lost. The greater the variance of E, relative to the variance of X, the less precise is X as a measure of T.

The degree of measurement error is not an inherent property of a biomarker test, but rather is a property of the test applied using a particular protocol to a specific population. Therefore, the error not only will vary between two assays which measure the same biomarker, but also can vary for a single test when applied using a different protocol or when applied to different population groups. Moreover, the validity coefficient is dependent on the variance of the true biomarker in the population (σ_T^2), so that even if the error variance σ_E^2 were the same for two populations, ρ_{TX} would differ if σ_T^2 differed. Therefore, a validity study done on one population may not directly apply to another study population. Finally, measurement error could differ between the study groups in the parent study, e.g. there could be differential error between cases and controls in a case–control study.

The terminology surrounding measurement error varies between fields. In this chapter, the terms validity, accuracy and measurement error are used as general terms reflecting the relationship between X and T, including both the concepts of bias and precision. In laboratory quality control, the terms validity and accuracy are sometimes used to refer to unbiasedness only.

Effects of error in a continuous biomarker on epidemiological studies

When the bias and validity coefficient of the biomarker (X) are known, one can estimate the impact of the degree of measurement error in X on the parent epidemiological study.

Several types of epidemiological study designs are discussed, and both differential and non-differential measurement errors are considered. First, however, the effect of measurement error on a single study population is discussed.

The effect of measurement error on the observable mean and variance

In a single study population, both the mean and variance of the measured biomarker X would differ from the true mean and variance due to measurement error. Using the above model, the population mean of X would differ from the true mean (the population mean of T) by b:

$$\mu_X = \mu_T + b$$

The population variance of X, based on the model and assumptions, would be (Allen & Yen, 1979):

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 = \sigma_T^2/\rho_{TX}^2 \quad (6)$$

Thus, the variance of X in the population is greater than the variance of T, due to the addition of the variance of the measurement error. For example, if the validity coefficient (ρ_{TX}) were 0.8, then the variance of measured X would be 56% greater than the variance of T ($\sigma_X^2 = \sigma_T^2/0.8^2 = 1.56\ \sigma_T^2$ by equation 6).

Figure 1 demonstrates the effect of measurement error on the distribution of X in a popula-

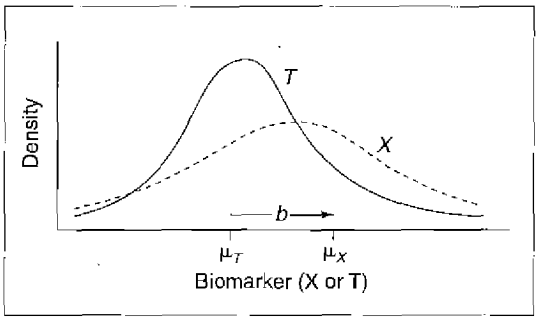

Figure 1. Distribution of true (T) and measured (X) biomarkers.

tion, using a normally distributed biomarker and normally distributed error as an example. The bias in the measure causes a shift in the distribution of X compared with T. The increased variance of X compared with T (measured by ρ_{TX}) causes a flattening of the distribution of X. Even if a measure were correct on average ($b = 0$), there could still be substantial measurement error due to lack of precision, which could lead to a greater dispersion in the measured exposures.

The effect of differential measurement error in a continuous biomarker in a two-group study

While measurement errors have an effect on the observable mean and variance of an exposure variable within a single population, a greater concern would be the effect of measurement errors when comparing a biomarker between two groups. This section and the next section cover epidemiological study designs in which the biomarker is compared between two groups. When the two groups being compared are exposure groups, e.g. a highly exposed group and an unexposed group, the common measure of association is the difference in the group means of the biomarker. When the groups being compared are a case group and a control group, the common measure of association between a biomarker and disease is the odds ratio, which is often expressed as the odds ratio of disease for one level of the biomarker versus another (usually lower) level, or as the odds ratio of disease for a u unit increase in the level of the biomarker. (These results do not apply to odds ratios expressed as odds of disease for the upper quantile of the biomarker versus the lowest quantile—see Armstrong *et al.*, 1992.)

For a two-group epidemiological study, the biomarker measure X_{1i} for the ith person in the first group (exposed or diseased group) differs from that person's true exposure T_{1i} by the systematic bias (b_1) in X within that group and by error in subject i's measure:

$$X_{1i} = T_{1i} + b_1 + E_{1i}$$

and, similarly, for the second group (unexposed or non-diseased group):

$$X_{2i} = T_{2i} + b_2 + E_{2i}$$

Differential measurement error occurs when b_1, the bias in the first group, differs from b_2, the bias in the second group, and/or the variance of E_1 differs from the variance of E_2.

Errors in the measurement of the biomarker X would affect the measure of the association in the epidemiological study. The effect of differential measurement error on the observable difference between groups in the biomarker means is:

$$\mu_{X_1} - \mu_{X_2} = (\mu_{T_1} - \mu_{T_2}) + (b_1 - b_2) \qquad (7)$$

This states that the observable difference between groups is equal to the true difference plus the difference in the biases of the biomarker test between groups. This equation holds even when the 'perfect' measure T is not perfectly precise, as long as T is unbiased (or at least not differentially biased between groups).

The effect of differential measurement error in X on the odds ratio can only be easily quantified when certain simplifying assumptions are made. Results can be derived for two-group studies under the following assumptions: (a) X_1 and X_2 are modeled as above with $\rho_{TE} = 0$ for each group, (b) T_1 and T_2 are normally distributed with means μ_{T_1} and μ_{T_2} respectively and the same variance, σ_T^2, and (c) E_1 and E_2 are normally distributed with mean zero and common variance, σ_E^2. The later assumption means that X_1 and X_2 are equally precise, so only differential bias is considered.

The above assumptions imply a logistic regression model for the probability of disease (pr(d)) as a function of true biomarker T, with a true logistic regression coefficient β_T (Wu *et al.*, 1986):

$$\log \frac{\text{pr(d)}}{\text{1-pr(d)}} = \alpha_T + \beta_T T$$

where

$$\beta_T = \frac{\mu_{T_1} - \mu_{T_2}}{\sigma_T^2}$$

The true odds ratio for any u unit increase in T would be $OR_T = e^{\beta_T u}$.

With measurement error in the biomarker test X, the assumptions also lead to a logistic model:

$$\log \frac{\text{pr(d)}}{\text{1-pr(d)}} = \alpha_O + \beta_O X,$$

where

$$\beta_O = \frac{(\mu_{T_1} - \mu_{T_2}) + (b_1 - b_2)}{\sigma_T^2 / \rho_{TX}^2}$$

The observable logistic regression coefficient, β_O, differs from β_T due to the measurement error in X. β_O can be expressed in terms of β_T (if $\beta_T \neq 0$) as follows (Armstrong *et al.*, 1989):

$$\beta_O = \left(1 + \frac{b_1 - b_2}{\mu_{T_1} - \mu_{T_2}}\right) \rho_{TX}^2 \beta_T \qquad (8)$$

Since $(b_1 - b_2)$ can be any magnitude and can be either positive or negative, the observable logistic regression coefficient, β_O, could be greater than, less than, or even have a different sign than, the true coefficient β_T. The observable odds ratio for any u unit increase in X, $OR_O = e^{\beta_O u}$, could be towards the null value of 1, away from the null or cross over the null value compared with the true odds ratio.

Figure 2 gives a graphical presentation of differential measurement error, in particular differential bias between cases and controls. In the figure, the true mean biomarker level in the diseased group, μ_{T_1}, is greater than the true mean biomarker level in the non-diseased group, μ_{T_2}. This would lead to a positive slope in the true odds ratio curve. In this example, the bias for the non-diseased group is positive, so the distribution of X_2 is shifted to the right relative to T_2, and the bias among those with disease is negative so that the distribution of X_1 is shifted to the left relative to T_1. This would lead the observable odds ratio curve to cross over

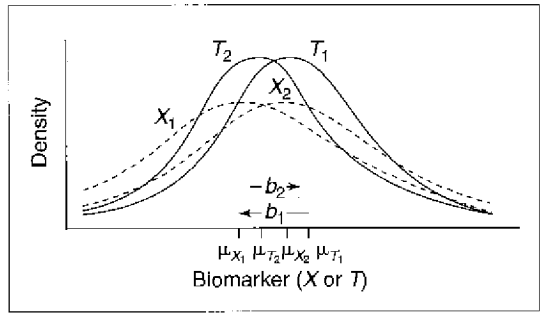

Figure 2. Effect of differential measurement error ($b_1 \neq b_2$) on distribution of true (T) versus measured (X) biomarker in a two-group study.

the null value of 1 (it would slope downwards from 1 as X gets larger, rather than upwards).

Differential measurement error should be a concern in a case–control study when the biomarker is measured within the pre-clinical disease phase before diagnosis, or any time after diagnosis, and the marker is not fixed. This concern is illustrated by several case–control studies that found low serum cholesterol at the time of diagnosis to be a risk factor for colon cancer (Law & Thompson, 1991), which could be an artefact if increased catabolism or other effects of colon cancer reduce serum cholesterol. As an example of the interpretation of equation (8), if true mean serum cholesterol for cases 10 years before diagnosis (before the effects of disease) was 6.0 nmol/l μ_{T_1}, while the true mean serum cholesterol among controls at a comparable time period was 5.8 μ_{T_2}, then the true odds ratio for colon cancer would be greater than 1, say $OR_T = 1.05$ for a 1 nmol/l increase ($\beta_T = \ln(1.05) = 0.05$). However, at the time of diagnosis (and a comparable time period for controls), there was substantial error in the measure of serum cholesterol (X) as a measure of T. Suppose the effect of the disease were to cause serum cholesterol at the time of diagnosis for cases (X_1) to be 0.5 nmol/l lower, on average, than 10 years earlier ($b_1 = \mu_{X_1} - \mu_{T_1} = -0.5$), while among controls the serum cholesterol increased over the 10 years by 0.5 nmol/l ($b_2 = \mu_{X_1} - \mu_{T_2} = 0.5$). Furthermore, suppose that $\rho_{TX} = 0.8$. Then, by equation (8);

$$\beta_O = \left[1 + \frac{-0.5 - 0.5}{6.0 - 5.8}\right] 0.8^2 \times 0.05$$

$$= -0.13$$

and

$$OR_O = e^{\beta_O} = 0.88$$

Under these assumptions, a true small increased risk ($OR_T = 1.05$) associated with serum cholesterol would appear as a decreased risk ($OR_O = 0.88$).

Differential bias is a greater concern than differential precision, because, as described above, differential bias can lead to a shift in the distribution of the biomarker in one group relative to the other. Differential measurement error will also occur if σ_E^2 differs between groups. If there were no differential bias but the σ_E^2-values differed (and σ_T^2-values were equal for the two groups), the shape of the odds ratio function could change. For example, the observable odds ratio curve could be U-shaped when the true exposure–disease relationship is increasing (Gregorio *et al.*, 1985).

The effect of non-differential measurement error in a continuous biomarker in a two-group study

When the assumptions given in the above section hold, there is non-differential misclassification when there is equal bias ($b_1 = b_2$) and equal error variance (or equivalently equal ρ_{TX}) in the biomarker test when applied to the two groups in the parent epidemiological study. Figure 3 illustrates the effects of non-differential misclassification. Under non-differential misclassification, the two distributions may shift, but they are not shifted with respect to each other, because there is equal bias for the two groups. Thus, based on the model presented, the observable difference in the mean values of X between groups is equal to the true difference:

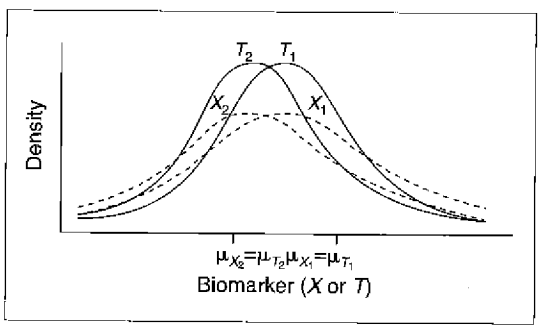

Figure 3. Effect of non-differential measurement error (equal bias and precision) on distribution of true (T) versus measured (X) biomarker in a two-group study.

$$(\mu_{X_1} - \mu_{X_2}) = (\mu_{T_1} - \mu_{T_2}) \qquad (9)$$

However, the lack of precision flattens each distribution and leads to more overlap and less distinction between the distributions of X_1 and X_2 compared with the true distributions. The variance of the observed difference between groups in the mean biomarker would be expected to increase by a factor $1/\rho_{TX}^2$.

Under non-differential misclassification, the odds ratio curve is flattened towards the horizontal line of odds ratio equal to 1 for all X. Equation (8) can be simplified to (Whittemore & Grosser, 1986; Wu *et al.*, 1986):

$$\beta_O = \rho_{TX}^2 \beta_T \qquad (10)$$

If $OR_T = e^{\beta_T u}$ is the true odds ratio for a u unit increase in T, and $OR_O = e^{\beta_O u}$ is the observable odds ratio for a u unit increase in X, then:

$$OR_O = OR_T^{\rho_{TX}^2} \qquad (11)$$

This states that the observable odds ratio for any fixed difference in units of the biomarker is equal to the true odds ratio for the same fixed difference to the power ρ_{TX}^2. Since $0 < \rho_{TX}^2 < 1$, the observable odds ratio will be closer to the null value of 1 (no association) than the true odds ratio. The observable odds ratio does not cross over the null value if X and T are, at a minimum, positively correlated.

While the difference in the bias of X between groups ($b_1 - b_2$) can play a major role in the bias in the odds ratio under differential measurement error (equation 8), equation (11) shows that the attenuation in the odds ratio under non-differential misclassification is a function of the precision of X (measured by ρ_{TX}), but not of the bias in X.

Prentice (1982) has shown that, under similar assumptions, equation (10) also applies approximately to estimates of β obtained from the proportional hazards model for data from cohort and matched case–control studies.

Examples of the effects of non-differential measurement error in a normally distributed biomarker on the odds ratio, based on the attenuation equation (11), are given in Table 2. The table shows that biomarkers with a validity coefficient ρ_{TX} of 0.5 would obscure all but the strongest associations.

For example, if the true odds ratio for a u unit change in the biomarker were 4.0, this would be attenuated to an observed odds ratio of 1.41. Furthermore, measures as precise as $\rho_{TX} = 0.9$ still lead to a modest attenuation—for example, a true odds ratio of 4.0 would be attenuated to 3.07.

The effect of independent measurement error in a study of the correlation between a continuous biomarker and a continuous variable

Often, the parent study which uses a biomarker will be one in which a continuous biomarker X will be compared with another continuous variable Y, e.g. a questionnaire measure of an exogenous exposure or another biomarker. Suppose both X and Y are measured with error, with the measurement error given by the model presented in 'the model of measurement error' section, with the true value of X denoted by T_X, the true value of Y denoted by T_Y, the subject error in X denoted by E_X and the subject error in Y denoted by E_Y. There is independent measurement error if E_X is uncorrelated with T_Y, E_Y is uncorrelated with T_X, and E_X is uncorrelated with E_Y. Under these assumptions, the observable correlation coefficient of X with Y is given by (Allen & Yen 1979):

$$\rho_{XY} = \rho_{T_XT_Y} \times \rho_{T_XX} \times \rho_{T_YY} \qquad (12)$$

This equation states that the observable correlation ρ_{XY} in the parent study is equal to the true correlation $\rho_{T_XT_Y}$ times the validity coefficient of X (ρ_{T_XX}) times the validity coefficient of Y(ρ_{T_YY}). Note that the bias in X and the bias in Y do not enter the equation. This is because adding a constant to X and Y does not change the correlation between X and Y.

As an example, suppose one were conducting a study of serum β-carotene (X) in relation to dietary β-carotene as measured by a food frequency questionnaire (Y). Assume that the true time period of interest is the prior year, but serum β-carotene has error primarily because it is only measured once during the year. The food frequency measure has error primarily because individuals cannot accurately recall diet over the past year. Suppose that the true correlation of serum β-carotene over the prior year (T_X) and dietary intake of β-carotene over the past year (T_Y) is 0.7. Assume that the correlation of a single measure of serum β-carotene (X) with 1 year average serum β-carotene (T_X) is 0.8, and the correlation of the food frequency estimate of dietary β-carotene (Y) with true dietary intake of β-carotene (T_Y) is 0.6. Also assume that the errors in X and Y are independent of true X and Y and of each other. Then the observable correlation coefficient between X and Y (ρ_{XY}) in the study would be (by equation 12) $0.7 \times 0.8 \times 0.6 = 0.34$, a considerable attenuation from the true association of 0.7.

Table 2. Effect of non-differential measurement error in a normally distributed biomarker X on the observable odds ratio (OR_O)

		True OR = 2.0	True OR = 4.0
ρ_{TX}	ρ_X	OR_O	OR_O
0.50	0.25	1.19	1.41
0.60	0.36	1.28	1.65
0.70	0.49	1.40	1.97
0.75	0.56	1.48	2.18
0.80	0.64	1.56	2.42
0.85	0.72	1.65	2.72
0.90	0.81	1.75	3.07
0.95	0.90	1.87	3.49

ρ_{TX} is the validity coefficient of X.
ρ_X is the reliability coefficient of X under the parallel test model (see text): $\rho_X = \rho_{TX}^2$.
The true OR is the odds ratio for a u unit difference in T.
OR_O is the observable odds ratio for a u unit difference in X.
Computed from equation (11). See text for model and assumptions.

Measuring the error in a continuous biomarker using a reliability study

The term reliability is generally used to refer to the reproducibility of a measure, i.e. how consistently a measurement can be repeated on the same subjects. Reliability can be assessed in a number of ways, but only one type will be covered in this chapter. Intramethod reliability studies measure the reproducibility of an instrument on the same subjects repeated two or more times using the identical method or with some variation. For example, a comparison could be made of a biomarker from a single specimen analysed in two

batches or by two laboratory technicians or from two specimens on each subject collected at two points in time. Reliability studies in which two different analytic methods are compared, with one better than the other but neither perfect (inter-method reliability studies) are not covered here (see Armstrong *et al.*, 1992). Measures of reliability are primarily important for what they reveal about the validity of a biomarker test, because the bias in the odds ratio in the parent epidemiological study is a function of the validity of the biomarker measure.

This section covers the interpretation of measures of reliability in terms of measures of validity for continuous biomarkers. The purpose is to provide some general concepts for the interpretation of reliability studies. Only continuous measures are covered because the relationship between the reliability and validity of categorical measures is more complex.

A model of reliability and measures of reliability

Suppose each person in a population of interest is measured two or more times using the same biomarker test to be used in the parent study. For a given subject *i*, two (or more) biomarker measurements, X_{i1} and X_{i2}, are obtained. (Note: in this section, X_1 and X_2 refer to the two measures per subject, not to two groups of subjects as in the earlier sections.) The simple measurement error model described above applies to each measure:

$$X_{i1} = T_i + b_1 + E_{i1}$$
$$X_{i2} = T_i + b_2 + E_{i2}$$

Both X_{i1} and X_{i2} are measures of the subject's true biomarker T_i, but with different errors. In a reliability study, information is available on X_1 and X_2 for each subject, but not on T. A reliability study can yield estimates of the mean of X_1 and X_2(μ_{X_1} and μ_{X_2}) and of the correlation between the two measures, $\rho_{T_XT_Y}$, termed the reliability coefficient.

The intraclass correlation coefficient is generally used as the reliability coefficient [see Fleiss (1986) or Armstrong *et al.* (1992) for formulas for its computation]. The intraclass correlation differs from the Pearson correlation coefficient in that it includes any systematic difference between X_1 and X_2 (i.e. any difference between b_1 and b_2) as part of the subject error E (the error that varies from subject to subject). The assumption is that in the parent epidemiological study, each subject will be measured once, by either X_1 or X_2 (e.g. either by laboratory technician 1 or 2). Therefore, any systematic difference between X_1 and X_2 would not be a systematic bias affecting everyone in the parent study, but would vary between subjects because some are measured by X_1 and some by X_2. Because X_1 and X_2 will be used as interchangeable measures of X in the parent study, and because more than two replicates per subject can be used to compute the intraclass correlation coefficient, the reliability coefficient of X can also be written as ρ_X.

Two measures of the validity of a continuous exposure measure were shown to be important in assessing the impact of measurement error: the bias and the validity coefficient. Unfortunately, reliability studies generally cannot provide information on the bias in X. The inability of many reliability study designs to yield information on bias, particularly on differential bias between study groups, is a major limitation. It should be recalled, however, that in the case of non-differential measurement error (and certain other assumptions), the attenuation equations depend only on the validity coefficient and not on the bias. The reliability coefficient does provide information about the validity coefficient, and thus can be used to estimate the effects of measurement error on the parent study under the assumption of non-differential measurement error.

Relationship between reliability and validity under the parallel test model

When certain assumptions are met, reliability studies can yield information about the validity coefficient. One such set of assumptions is the model of parallel tests (Lord & Novick, 1968; Nunnally, 1978; Allen & Yen, 1979; Bohrnstedt, 1983). The first assumption of the parallel test model is that the error variables, E_1 and E_2, are not correlated with the true value T. It is further assumed that E_1 and E_2 have equal variance, σ_T^2. This also implies that X_1 and X_2 have equal variance and that X_1 and X_2 are equally precise ($\rho_{TX_1} = \rho_{TX_2}$). This is usually a reasonable assumption in intramethod reliability studies, since X_1 and X_2 are measurements from the same instrument. Finally, it is assumed that E_1 is not correlated with E_2. This important (and restrictive) assumption implies, for example, that an individual who has a positive error, E_1, on the first measurement is equally likely to have a positive

or a negative error, E_2, on the second measurement. These assumptions are often summarized by saying that two measures are parallel measures of T if their errors are equal and uncorrelated.

Under the assumption of parallel tests, it can be shown that (Allen & Yen, 1979):

$$\rho_X = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2} = \rho_{TX}^2 \qquad (13)$$

or equivalently

$$\rho_{TX} = \sqrt{\rho_X} \qquad (14)$$

These equations state that the reliability coefficient, ρ_X, is equal to the square of the validity coefficient of X, ρ_{TX}. This result is important, because it shows that if the assumptions are correct, the reliability coefficient, which is a measure of the correlation between two imperfect measures, can be used to estimate the correlation between T and X, without having a perfect measure of T. The correlation of the replicates of X is less than the correlation of X with T because each replicate has measurement error.

A reliability study of a biomarker test can often be assumed to have equal and uncorrelated errors if (1) the replicates are sampled over the entire time period to which the true biomarker is intended to relate; (2) the specimen handling, storage and analytical techniques vary in the reliability study as they will in the parent study; and (3) the true exposure is defined as the mean measure over the relevant time period of repeated measures of the assay.

A study by Toniolo *et al.* (1994), which examined the reliability of serum hormone levels in 77 postmenopausal women, provides a good example. Selected women in a prospective study of serum estrogens and breast cancer risk who had blood drawn 2–3 times over a 1–2 year period were included in the reliability study. The blood was stored and analysed as in the parent study. The reliability coefficient (intraclass correlation coefficient) was 0.51 for total estradiol and 0.77 for percentage unbound estradiol. The repeated measures in this study are close to parallel test model: the errors on each of the repeated measures can be assumed to be equal because the same test procedure was repeated, and the errors are likely to be independent (i.e. a woman whose hormone measure was higher than her 'true' average value on one measure is not more likely to be higher than her true value on another measure). This study also measures most sources of error—blood processing, storage, random laboratory error (within-batch error) and variation due to changes in day-to-day and long-term variations of plasma hormones within women. Thus, the estimated validity coefficient (ρ_{TX}) for a single measure of total estradiol (X) as a measure of average estradiol over 1–2 years (T), based on equation (14), is 0.71, and it is 0.88 for percentage unbound estradiol.

Based on equation (13), the results in the last section on the effects of measurement error could have been (and often are) expressed in terms of ρ_X rather than ρ_{TX}^2. For example, equation (11) can be written:

$$OR_Q = OR_T^{\rho_X} \qquad (15)$$

Examples of the bias in the odds ratio from various degrees of unreliability are given in Table 2. For example, a biomarker with a reliability coefficient of 0.64 (under the parallel test model) would attenuate a true odds ratio of 4.0 to an observed odds ratio of 2.4. When ρ_X is substituted for ρ_{TX}^2 in the attenuation equation, the equation applies only when the meaning of the reliability coefficient is restricted to the correlation between parallel measures of T. However, the term 'reliability coefficient' is often used to refer to the correlation between repeated measures, ρ_X, even when the assumptions of parallel tests do not hold.

Relationship between reliability and validity when the errors are correlated

In real reliability studies, the assumptions of parallel tests are often incorrect. One assumption of the model of parallel tests that is often violated is the assumption of uncorrelated errors. Often, $\rho_{E_1E_2} > 0$ In other words, the error in one measure is positively correlated with the error in the other. Correlated errors occur when the sources of error in the first measurement on a subject tend to repeat themselves in the second. For example, if, in a reliability study, blood was drawn once on each subject and analysed twice in different batches, and the true marker of interest were mean β-carotene for the 2 years surrounding the time of measurement, there would be correlated error. This is because a person whose β-carotene level on the

first measure was higher than their true mean level (perhaps due to a seasonal variation in intake of β-carotene) would also be likely to have a β-carotene level that was higher than their true value on the second measure, because the second measure used the same specimen. The errors are correlated because part of the error is repeated in both X_1 and X_2.

To interpret a reliability study, one should evaluate whether there are potential sources of correlated errors between the two measures. As outlined in Box 1, there is a wide range of sources of measurement error, and most of these could be sources of correlated errors.

When the errors of the measures in a reliability study are positively correlated, then the reliability study can only yield an upper limit for the validity coefficient. Specifically, when X_1 and X_2 are equally precise and the assumptions of the above model hold (except $\rho_{E_1E_2} > 0$), then the validity coefficient is less than the square root of the reliability coefficient (Walker & Blettner, 1985):

$$\rho_{TX} < \sqrt{\rho_X} \qquad (16)$$

Thus, a measure can be reliable (repeatable) even if it has poor validity. While a low reliability coefficient implies poor validity, a high reliability does not necessarily imply a high validity coefficient. The high reliability may be due instead to repeated errors within subjects. The reliability coefficient is only diminished by part of the error in X (the part that is not repeated in X_1 and X_2), whereas the validity coefficient is a measure of all sources of error.

Reliability studies often need to be more complex than can be described here. For example, in the study by Toniolo *et al.* (1994) described above, the variability due to batch effects was removed from the reliability coefficient. This is because, in the parent study, each matched set of cases and controls were to be analysed within a batch so that batch differ-

Table 3. Equations for the true measure of association as a function of the observable measure of association and the biomarker measurement error.[a]

Equation	From equation	Differential or non-differential
Binary biomarker, two-group comparison		
True difference = $P_1 - P_2$ $OR_T = \frac{P_1(1-P_2)}{P_2(1-P_1)}$ where $P_1 = (p_1 - 1 + spec_1)/(sens_1 + spec_1 - 1)$ and $P_2 = (p_2 - 1 + spec_2)/(sens_2 + spec_2 - 1)$	(1–3)	Differential or non-differential
$P_1 - P_2 = (p_1 - p_2)/(sens + spec - 1)$	(4)	Non-differential
Continuous biomarker, two-group comparison		
$\mu_{T_1} - \mu_{T_2} = (\mu_{X_1} - \mu_{X_2}) - (b_1 - b_2)$	(7)	Differential
$OR_T = e^{\beta_T}$ where $\beta_T = \left(1 - \frac{b_1 - b_2}{\mu_{X_1} - \mu_{X_2}}\right)\frac{\beta_O}{\rho^2_{TX}}$	(8)	Differential
$\mu_{T_1} - \mu_{T_2} = \mu_{X_1} - \mu_{X_2}$	(9)	Non-differential
$OR_T = OR_O^{1/\rho^2_{TX}}$	(11)	Non-differential
$OR_T = OR_O^{1/\rho_X}$	(15)	Non-differential
Continuous biomarker, continuous outcome		
$\rho_{T_XT_Y} = \rho_{XY}/(\rho_{T_XX} \times \rho_{T_YY})$	(12)	Non-differential

[a]See text for the notation, the assumptions used in derivation of equations and the interpretation of equations.

ences would not contribute to measurement error.

The coefficient of variation, CV, provides only limited information about measurement error because the CV is the ratio of σ_E to $\bar{X}$, but it is the ratio of σ_E to σ_X (as in equation 13 above) that is needed to understand the impact of measurement error.

Adjusting the results of epidemiologic studies for biomarker error

The emphasis of this chapter has been an explanation of how the observable odds ratio or other measure of association in the parent epidemiological study is a function of the measurement error in the biomarker and the true measure of association. However, the same models can be applied in the opposite direction. By use of estimates of the biomarker measurement error from a validity or reliability study, it is possible to adjust the observed measure of association from the parent epidemiological study to yield an estimate of the true association. However, because the assumptions used in the models cannot be known to be correct (see below), the results of the adjustment equations should not be considered to be the 'true association', but rather an indicator of the degree of bias in the observed odds ratio or other measure of association.

These 'adjustment' equations and the equations from which they were derived are given in Table 3. Because under non-differential measurement error the observable association is attenuated from the true association, these adjustment equations are called 'deattentuation' equations when non-differential measurement error is assumed.

Data from a study by Winawer *et al.* (1990) are adapted as an example of adjusting for the effects of differential measurement error in a two-group study of a continuous biomarker. In this study, serum cholesterol at the time of diagnosis (X) was compared between 43 cases with colon cancer and 43 controls (see Table 4).

The observed case–control difference in serum cholesterol was:

$$\bar{X}_1 - \bar{X}_2 = 5.56 - 6.47 = -0.91$$

Thirty-two subjects (16 cases and 16 controls) had serum cholesterol measures available from 10 years prior to diagnosis, which can serve as T because it is unlikely to be biased. (As noted earlier, the bias in X can be estimated by comparison with T if T is unbiased, even if T is not perfectly precise.) A comparison of X with T for these 32 subjects could serve as a validity study to estimate bias. The bias in X among the cases could be estimated from the validity study (with $\bar{X}$, $\bar{T}$ serving as estimates of μ_X, μ_T) as:

$$\hat{b}_1 = \bar{X}_1 - \bar{T}_1 = 5.75 - 6.21 = -0.46,$$

i.e. serum cholesterol at the time of diagnosis underestimates true serum cholesterol by 0.46 among cases. The bias in X among the controls could be estimated from the validity study as:

$$\hat{b}_2 = \bar{X}_2 - \bar{T}_2 = 6.30 - 5.72 = 0.58,$$

i.e. serum cholesterol (at time of diagnosis of cases) overestimates true serum cholesterol by 0.58. Clearly, there is differential bias.

To correct the observed difference in serum cholesterol in the parent study for the effects of this differential bias, the correction equation from Table 3 based on equation (7) is used, with $\bar{X}$ and $\bar{T}$ serving as estimates of μ_X and μ_T:

$$\widehat{(\mu_{T_1} - \mu_{T_2})} = (-0.91) - (-0.46 - 0.58) = 0.13$$

This suggests that there is little difference between colon cancer cases in 'true' (before disease) serum cholesterol.

As an example of deattenuation under non-differential measurement error, suppose a cohort

Table 4. Example of adjustment for effects of differential measurement error

	Serum cholesterol (nmol/l)	
	$\bar{X}$	$\bar{T}$
Case–control study (parent study)		
Cases (n = 43)	5.56	
Controls (n = 43)	6.47	
Validity study		
Cases (n = 16)	5.75[a]	6.21
Controls (n = 16)	6.30[a]	5.72

[a]Data created to serve as an example.

study of occupational exposure to chlorophenates, assessed by urinary chlorophenate concentration, in relation to soft tissue sarcoma yielded an odds ratio of 1.3 for each 100 μg/L increase in concentration. Suppose also that a validity study among a subset of the subjects yielded an estimate of 0.6 for the validity coefficient, ρ_{TX}, between urinary cholorophenate concentration and industrial records of exposure (assumed here to be a near-perfect measure). Then, information from these two studies could be used to adjust the observed odds ratio to yield an estimate of the true odds ratio (based on the equation in Table 3 derived from equation (11), provided the assumptions hold):

$$OR_T = (1.3)^{1/0.6^2} = 2.1$$

This suggests that biomarker measurement error may have led to the weak observed association between the biomarker and the disease, because the observed odds ratio is consistent with a true odds ratio of 2.1 for each 100 μg/l increase in chlorophenate concentration.

Information from reliability studies can also be used in adjustment procedures, to the extent that the reliability study provides information about the validity of the exposure variable. If the reliability study were of two parallel measures, ρ_X could be substituted for ρ^2_{TX} as in the equation in Table 3 derived from equation (15). As an example, the prospective study by Toniolo *et al.* (1995) of serum hormone levels and breast cancer risk among 7063 postmenopausal women (among whom 130 developed breast cancer), observed an odds ratio of 1.8 for total estradiol >44 pg/ml (versus <20) and an odds ratio of 2.0 for percentage unbound estradiol >1.53% (versus <1.20%). These odds ratios can be corrected for the effect of measurement error, based on the reliability coefficients of 0.51 and 0.77 for total estradiol and percentage unbound estradiol, respectively, from the reliability study described in an earlier example. Using equation (15), the estimated 'true' odds ratio is 3.2 for total estradiol >44 pg/ml, and 2.5 for percentage unbound >1.53%. This suggests that total estradiol might be strongly related to breast cancer risk, but its effect is obscured by the error in its measurement.

Reliability studies yield only an upper limit for the validity coefficient when there are correlated errors in the repeated measures. Then, if equation (15) is applied, this leads to a conservative (closer to the null) estimate of the true odds ratio under non-differential measurement error. In other words, the estimate would only be 'deattenuated' for the random part of measurement error, and not for the part that was repeated across measures in the reliability study.

While adjustment procedures may aid in understanding the results of a study, caution should be exercised in interpreting these results. First, the assumptions used in the derivation of the equations in Table 3 may not be appropriate. In particular, an assumption of non-differential measurement error could be incorrect, so it is preferable if the biomarker measurement error can be assessed separately for the two study groups to account for differential misclassification. For continuous exposures, the assumptions of the simple measurement error model, normality of the biomarker and its error, and the logistic model of disease–biomarker relationship often also fail to hold. Second, both the observed measure of association between the biomarker and outcome and the estimated measurement error have sampling errors; this needs to be considered in estimating the true association. Third, the estimates of the measurement error should be estimates from the same population(s) as the parent study to be corrected, yet such estimates may not be available. Finally, the presence of covariates modifies the effect of biomarker measurement error. Information on the multivariate measurement error structure of the biomarker and covariates is required in order to correct fully for measurement error. Therefore, unless these issues have been accounted for, the emphasis of the adjustment procedure should be on interpretation of the observed estimate of effect, and not on the corrected estimates.

A great deal of work has been done on adjustment procedures, or, more generally, procedures that incorporate information from a validity or reliability study into the statistical analysis of the disease–exposure relationship (Tenebein, 1970; Barron, 1977; Copeland *et al.*, 1977; Prentice, 1982; Greenland & Kleinbaum, 1983; Clayton, 1985; Stefanski & Carroll, 1985; Whittemore & Grosser, 1986; Espeland & Hui, 1987; Fuller, 1987; Armstrong *et al.*, 1989; Rosner *et al.*, 1989; Qizilbash *et al.*, 1991; Pepe & Fleming, 1991; Carroll *et al.*, 1995). Many of these procedures take into consid-

eration some of the issues discussed above; in particular, some make less restrictive assumptions about the error model, yield confidence intervals that incorporate sampling error from the reliability study, and/or allow a multivariate measurement error structure. These methods may prove to be useful in accounting for measurement error.

Summary

Before embarking on an epidemiological study which uses a biomarker, it is extremely important to understand the measurement error in the biomarker. For a binary biomarker, the validity (the measurement error in a population) of the biomarker is quantified by its sensitivity and specificity. For a continuous biomarker, X, the validity can be estimated by the bias $\bar{X} - \bar{T}$ and by the validity coefficient, ρ_{XT} (correlation coefficient of X with T), a measure of precision. These estimates of measurement error ideally utilize a true measure of the biomarker (T) which has no sources of error and which integrates the biomarker over the time period of etiological interest (generally years). To assess whether the error is differential between cases and controls, separate studies on a group of cases and a group of controls are needed.

The researcher's first concern should be to rule out, as far as is practicable, the possibility of differential measurement error. Because differential measurement error can bias the odds ratio in any direction, the presence of differential error in a biomarker can invalidate the epidemiological study. Differential measurement error is a particular concern in case–control studies (and among the early cases in cohort studies) when the biomarker is not a fixed marker (e.g. genotype) and therefore could be influenced by pre-clinical disease, by the physical or emotional effects of the disease after diagnosis, and/or by treatment. Assessment of differential error requires pre-diagnostic specimens on a sample of cases and comparable early specimens on controls. These serve as the 'true' markers (without differential bias) for computation of sensitivity and specificity separately for cases and controls. For continuous variables, differential bias has the most untoward effects; this can be estimated by comparing $\bar{X} - \bar{T}$ for cases with $\bar{X} - \bar{T}$ among controls. Whether the degree of differential error is acceptable can be determined by estimating the effects of the error on the odds ratio (see equations (1) and (2) for binary markers and equation (8) for continuous biomarkers). Generally, very little differential error is acceptable.

When differential error is unlikely to be a problem, the researchers should focus on assessment of the non-differential error, or at least some of the major components of error in the biomarker. To assess the total error (the laboratory variation, variation from specimen collection and storage, and variation from short-, medium- and long-term biological variability), one would ideally conduct a validity study in which the biomarker to be used is compared with a perfect (true) measure. Then, the effect of non-differential measurement error on the odds ratio can be estimated by equations (1) and (3) for binary markers and equation (11) for continuous markers. For continuous markers (assuming a simple measurement error model), the effect of non-differential measurement error depends only on the validity coefficient and not on the bias (see equation 11).

Because validity studies are rarely possible, it is important to understand that one can measure much of the total error in a well-designed reliability study. Much of the error can be measured if one collects and analyses two (or more) specimens on a group of subjects in such a way that the error in one estimate is not repeated on another, e.g. the two specimens are collected at different times over the relevant etiological time period and are handled, stored and analysed with the same degree of variation (different specimen collectors/laboratory technicians/batches) as would occur in the parent epidemiological study. Reliability studies are generally analysed by the kappa coefficient for binary variables and by the intraclass correlation coefficient for continuous variables (see Fleiss, 1981, 1986; Armstrong *et al.*, 1992). The intraclass coefficient provides information about the validity coefficient (see equations 14 and 16) and therefore provides information about the degree of attenuation of the odds ratio due to non-differential measurement error. The degree of acceptable measurement error would depend on the magnitude of the true odds ratio (see equation 15), but generally an intraclass correlation (including all sources of errors) of less than 0.5 would not be acceptable.

Whether or not the total error can be studied, one should attempt to estimate some or all of the components of error. The epidemiologist and

laboratory scientists should develop studies of the laboratory error and other components of error. To assess the laboratory error, a blinded test–retest reliability study on split samples from a single specimen from each of a group of subjects, analysed in separate batches and by different laboratory technicians (in a way that reflects the design of the parent epidemiological study), would yield an intraclass correlation coefficient that measures the laboratory component of error. Similarly, other reliability studies could be designed to test the effect of handling, storage and short-, medium- and long-term biological variation. Epidemiologists should not assume that a small laboratory error implies that a measure is good, because these later sources of error introduced by the design and needs of the epidemiological study could be far greater than the laboratory component of the error. Often, multiple sources of error can be partitioned in a nested study design (see Dunn, 1989). When only some components of error are measured, the resulting intraclass correlation only provides an upper estimate of the validity coefficient (see equation 16). However, by estimating the components of error, the researcher can seek to improve those aspects having the most adverse effects. For example, enhanced laboratory quality control procedures would reduce laboratory error, or use of multiple specimens (over time) per subject would reduce the error due to short- or medium-term biological variation (see Armstrong *et al.*, 1992).

Finally, after the parent epidemiological study has been completed, estimates of biomarker measurement error can be used to adjust the results of the parent study for the effects of measurement error (or for those components of error that have been measured) (see Table 3). Cautious use of these methods can add insight into the bias caused by biomarker measurement error on the results of an epidemiological study.

References

Armstrong, B.G., Whittemore, A.S. & Howe, G.R. (1989) Analysis of case–control data with covariate measurement error: application to diet and colon cancer. *Stat. Med.*, 8, 1151–1163

Armstrong, B.K., White, E. & Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology*, Oxford, Oxford University Press, pp. 49–136

Barron, B.A. (1977) The effects of misclassification on the estimation of relative risk. *Biometrics*, 33, 414–418

Bohrnstedt, G.W. (1983) Measurement. In: Rossi, P.H., Wright, J.D. & Anderson, A.B., eds, *Handbook of Survey Research*, Orlando, FL, Academic Press, pp. 70–121.

Bross, I. (1954) Misclassification in 2 × 2 tables. *Biometrics*, 10, 478–486

Carroll, R.J., Ruppert, D. Stefanski, L.A. (1995) *Measurement Error in Nonlinear Models*, London, Chapman & Hall

Clayton, D. (1985) Using test–retest reliability data to improve estimates of relative risk: an application of latent class analysis. *Stat. Med.*, 4, 445–455

Cochran, W.G. (1968) Errors of measurement in statistics. *Technometrics*, 10, 637–666

Copeland, K.T., Checkoway, J., McMichael, A.J. & Holbrook, R.H. (1977) Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.*, 105, 488–495

de Klerk, N.H., English, D.R. & Armstrong, B.K. (1989) A review of the effects of random measurement error on relative risk estimates in epidemiologic studies. *Int. J. Epidemiol.*, 18, 705–712

Dunn, G. (1989) *Design and Analysis of Reliability Studies*, New York, Oxford University Press

Espeland, M.A. & Hui, S.L. (1987) A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*, 43, 1001–1012

Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions*, 2nd edn, New York, Wiley, pp. 188–236

Fleiss, J.L. (1986) *The Design and Analysis of Clinical Experiments*, New York, Wiley

Fuller, W.A. (1987) *Measurement Error Models*, New York, Wiley

Goldberg, J.D. (1975) The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. *J. Am. Stat. Assoc.*, 70, 561–567

Greenland, S. & Kleinbaum, D.G. (1983) Correcting for misclassification in two way tables and matched-pair studies. *Int. J. Epidemiol.*, 12, 93–97

Gregorio, D.I., Marshall, J.R. & Zielenzyny, M. (1985) Fluctuations in odds ratios due to variance differences in case–control studies. *Am. J. Epidemiol.*, 121, 767–774

Gullen, W.H., Bearman, J.E. & Johnson, E.A. (1968) Effects of misclassification in epidemiologic studies. *Public Health Reports*, 83, 914–918

Hansen, M.H., Hurwitz, W.N. & Bershad, M. (1961)

Measurement errors in censuses and surveys. *Bull. Int. Stat. Inst.*, 38, 359–374

Kleinbaum, D.G., Kupper, L.L. & Morgenstern, H. (1982) *Epidemiologic Research*, Belmont, CA, Lifetime Learning Publications, pp. 183–193, 220–241

Kotz, S. & Johnson, N.L. (1988) *Encyclopedia of Statistical Sciences*, vol. 9, New York, Wiley, pp. 662–663

Law, M.R. & Thompson, S.G. (1991) Low serum cholesterol and the risk of cancer: an analysis of the published prospective studies. *Cancer Causes Control*, 2, 253–261

Lord, F.M. & Novick, M.R. (1968) *Statistical Theories of Mental Test Scores*, Reading, MA, Addison-Wesley

McKeown-Eyssen, G.E. & Tibshirani, R. (1994) Implications of measurement error in exposure for the sample sizes of case–control studies. *Am. J. Epidemiol.*, 139, 415–421

Newell, D.J. (1962) Errors in the interpretation of errors in epidemiology. *Am. J. Publ. Health*, 52, 1925–1928

Nunnally, J.C. (1978) *Psychometric theory* (2nd edn). New York, McGraw-Hill, pp. 190–225

Prentice, R.L. (1982) Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69, 331–342

Qizilbash, N., Duffy, S.W. & Rohan, T.E. (1991) Repeat measurement of case–control data: correcting risk estimates for misclassification due to regression dilution of lipids in transient ischemic attacks and minor ischemic strokes. *Am. J. Epidemiol.*, 133, 832–838

Rosner, B., Willett, W.C. & Spiegelman, D. (1989) Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat. Med.*, 8, 1051–1069

Schiffman, M.H. & Schatzkin, A. (1994) Test reliability is critically important to molecular epidemiology: An example from studies of human papilloma infection and cervical neoplasia. *Cancer Res.*, 54, 1944s–1947s

Stefanski, L.A. & Carroll, R.J. (1985) Covariate measurement error in logistic regression. *Ann. Stat.*, 13, 1335–1351

Tenenbein, A. (1970) A doubling sampling scheme for estimating from binomial data with misclassifiction. J. *Am. Stat. Assoc.*, 65, 1350–1361

Toniolo, P., Koenig, K.L., Pasternack, B.S., Banerjee, S., Rosenberg, C., Shore, C.E., Strax, P. & Levitz, M. (1994) Reliability of measurements of total, protein bound and unbound estradiol in serum. *Cancer Epidemiol. Biomarkers Prev.*, 3, 47–50

Toniolo, P.G., Levitz, M., Zeleniuch-Jacquotte, A., Banerjee, S., Koenig, K.L., Shore, R.E, Strax, P. & Pasternack, B.S. (1995) A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J. Natl Cancer Inst.*, 87, 190–197

Walter, A.M. & Blettner, M. (1985) Comparing imperfect measures of exposure. *Am. J. Epidemiol.*, 121, 783–90

White, E., Kushi, L. & Pepe, M. (1994) The effect of exposure variance and exposure measurement error on study sample size: implications for the design of epidemiologic studies. *J. Clinical Epidemiol.*, 47, 873–880

Whittemore, A.S. & Grosser, S. (1986) Regression methods for data with incomplete covariates. In: Moolgavkar, S.H. & Prentice, R.L, eds, *Modern statistical methods in chronic disease*. New York, Wiley, pp. 19–34

Winawer, S.J., Fiehinger, B.J., Buchalter, J., Herbert, E. & Shike, M. (1990) Declining serum cholesterol levels prior to diagnosis of colon cancer. *J. Am. Med. Assoc.*, 263, 2083–2085

Wu, M.L., Whittemore, A.S. & Jung, D.L. (1986) Errors in reported dietary intakes: I. Short-term recall. *Am. J. Epidemiol.*, 124, 826–835

E. White
Cancer Prevention Research Program,
Fred Hutchinson Cancer Research Center,
Seattle, WA 98109, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Markers of internal dose: chemical agents

D. Coggon and M.D. Friesen

Biomarkers of internal dose measure the level of a carcinogen or one of its metabolites in a tissue or a body fluid such as urine or blood. The choice of a biomarker of internal dose for a particular epidemiological study or type of study requires careful consideration of the period of exposure to which the biomarker relates, host factors related to carcinogen metabolism, invasiveness of sampling, reliability and cost of the biomarker. Before a new biomarker is adopted, it is important to assess these characteristics in transitional studies to ensure that the biomarker will be applied appropriately. Biomarkers of internal dose have been applied most successfully in ecological studies and nested case–control studies, and are especially useful when they provide information about long-term carcinogen exposure.

Human exposure to carcinogens can be estimated by measuring available levels in food, water or air. Air levels, for example, have been widely used in industrial hygiene to asses industrial exposure to carcinogens. Measuring the external exposure, however, does not necessarily provide accurate information about the amount of carcinogen actually absorbed by an individual, a parameter which is of great importance to epidemiologists. This parameter, which has been called the 'internal dose', reflects total exposure by all routes (inhalation, ingestion and dermal absorption) and is influenced not only by the level of carcinogen in external media (air, food, water, etc.) but also by factors such as the amount of air breathed or food consumed by the individual during the period of exposure and the efficiency of absorption from the lung or gut.

This chapter will examine the use of biomarkers of internal dose to chemical agents. Other types of biomarkers, which measure the biologically effective dose or early biological effects of the carcinogen, are also currently available or under development. Several excellent books have reviewed the use of all these kinds of biomarkers of human carcinogen exposure (Hulka *et al.*, 1990; Groopman & Skipper, 1991; Armstrong *et al.*, 1992; Schulte & Perera, 1993).

Biomarkers of internal dose

Biomarkers of internal dose measure the level of the carcinogen or one of its metabolites in a tissue or body fluid such as urine or blood. Table 1 gives examples of how biomarkers of internal dose have been applied in epidemiological studies. The applicability of a biomarker of internal dose to cancer epidemiology depends on certain features which need to be characterized in transitional studies (Hulka, 1991) before application in the field.

To what period of exposure does the biomarker relate?

Although biomarkers of internal dose generally give information about recent exposure, the half-life of a carcinogen or its metabolite in human body fluids can vary from less than an hour to more than 10 years (Fig. 1). For example, the level of urinary methylhippuric acid, which has a half-life of only a few hours, is useful in occupational settings as a biomarker of recent exposure to xylene (Inoue *et al.*, 1993; Huang *et al.*, 1994), while urinary cadmium, which is related to body burden with an excretion half-life of over 10 years, is a useful long-term biomarker of exposure to cadmium (Ghezzi *et al.*, 1985). Biomarkers of internal dose are being increasingly used in the field of industrial hygiene, where validated biomarkers for measuring occupational exposures to a wide range of carcinogens in blood, urine or breath have been proposed by national industrial hygiene authorities (Table 2).

In general, it is past or long-term exposure that is of interest in cancer epidemiology; however, a biomarker that reflects recent exposure may still be of importance and provide information superior to that obtainable with traditional methods of exposure assessment. For habitual or repeated exposures

Table 1. Examples of biomarkers of internal dose applied in epidemiological studies

Biomarker	Body fluid	Exposure	Target organ	Reference
Aflatoxin B_1 and metabolites	Urine	Aflatoxin in the diet	Liver	Ross *et al.*, 1992; Qian *et al.*, 1994
Dioxins and furans	Blood fat	Herbicide production	All organs	Flesch Janys *et al.*, 1995
Lead	Blood	Lead battery production	Gastrointestinal tract	Gerhardsson *et al.*, 1995
N-nitrosamino acids	Urine	Endogenous formation	Oesophagus	Wu *et al.*, 1993
Arsenic	Urine	Copper smelting	Lung	Enterline & Marsh, 1982
Cotinine	Urine	Tobacco smoking	Lung	De Waard *et al.*, 1995
DDE	Serum	DDT	Breast	Wolff *et al.*, 1993; Krieger *et al.*, 1994

such as smoking or mycotoxins in the diet, for example, a small number of measurements with a biomarker of recent exposure may give a reasonable measure of longer-term exposure (Ross *et al.*, 1992; Qian *et al.*, 1994; De Waard *et al.*, 1995). The use of such biomarkers of recent exposure to asses longer-term exposure requires transitional studies to determine the appropriate number of samples and frequency of sampling.

How are biomarker levels affected by metabolic host factors?

The level of biomarker measured in a tissue or body fluid depends on metabolic factors specific to the individual under examination. These include how the carcinogen and its metabolites are partitioned in elimination or storage compartments such as urine or adipose tissue, individual levels of phase I and phase II enzymes, and the influence on these enzyme levels of other substances to which the subject is exposed.

For carcinogens occurring in food, where the intake in a normal diet can be accurately determined, it is feasible to validate exposure biomarkers in studies with human volunteers. This approach is being used in transitional studies to characterize interindividual differences in the urinary excretion of 1-hydroxypyrene-glucuronide (Kang *et al.*, 1995), a metabolite of polycyclic aromatic hydrocarbons (PAH), and the heterocyclic aromatic amines 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) and 2-amino-3,4-dimethylimidazo[4,5-*f*]-quinoxaline (MeIQx) (Sinha *et al.*, 1994) after consumption of charbroiled or pan-fried beef by human volunteers.

The situation is more complex in the case of agents to which exposure occurs by multiple routes and at levels that are difficult to measure. One example of such a complex situation is exposure to PAH, for which urinary excretion of 1-hydroxypyrene is a marker (Sherson *et al.*, 1992; Santella *et al.*, 1993).

Is sampling invasive?

Samples of urine and breath are fairly easy to obtain, and in adults, blood sampling does not usually pose ethical difficulties or cause major problems with compliance. On the other hand, some tissues may be more difficult to sample. For example, although the level of DDT accumulated in adipose tissue is an excellent measure of long-term exposure to the pesticide, it would be hard to obtain samples from large numbers of people who were not already undergoing surgery. However, blood levels of DDE, a metabolite of DDT, have been used to measure long-term exposure to DDT (Wolff *et al.*, 1993; Krieger *et al.*, 1994).

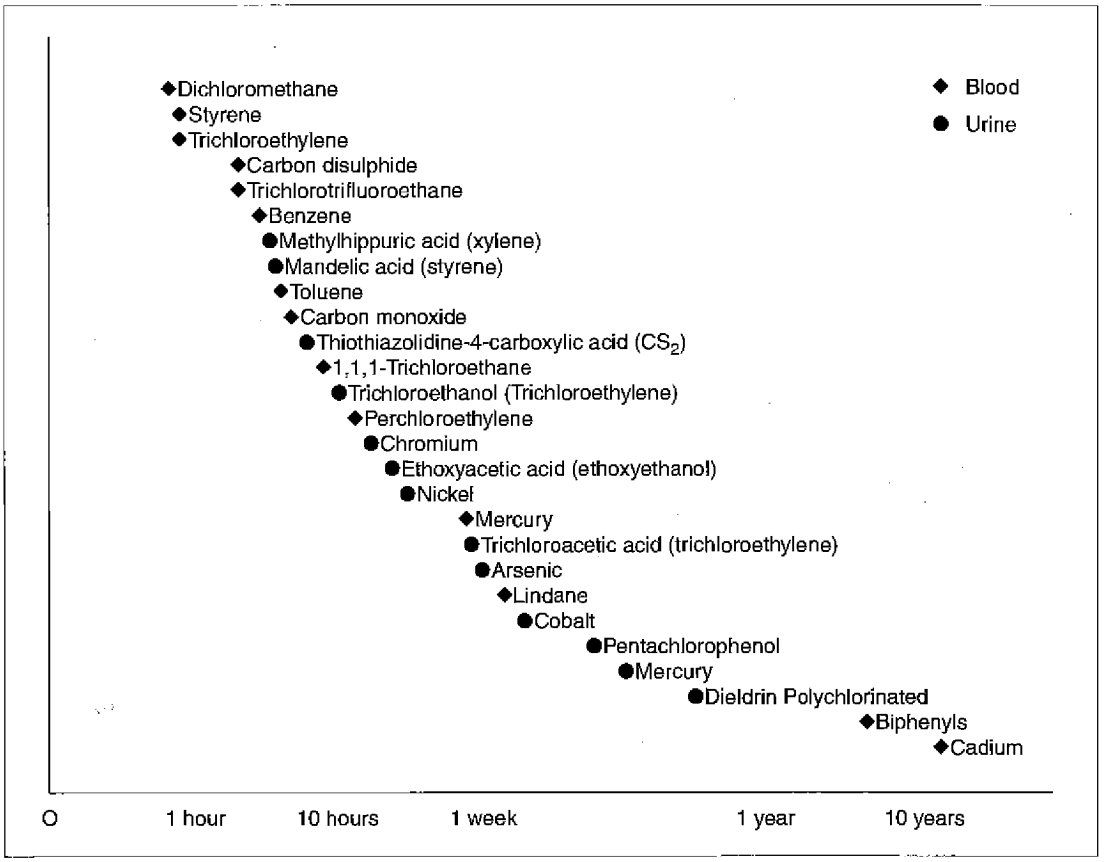

Figure 1. Excretion half-life of various biomarkers of internal dose in blood and urine (from data published by the UK Health and Safety Executive).

How much does the assay cost?

Many assays for measurement of internal dose, such as radioimmunoassay for the measurement of urinary cotinine as an index of exposure to environmental tobacco smoke (De Waard *et al.*, 1995), can be carried out on a large scale at a relatively low cost. Other markers such as measurement of dioxins in blood fat (Flesch Janys *et al.*, 1995) are still very expensive and thus applicable only to small numbers of samples, perhaps in a subset of a cohort under study.

Use of biomarkers of internal dose in epidemiological studies

Biomarkers of internal dose should be evaluated within the framework of epidemiological study design (Rothman *et al.*, 1995).

Cohort studies relating exposure to cancer.

Because cancers occur relatively infrequently, cohort studies require large numbers of subjects. For use in a standard cohort study, biomarkers of internal dose can be a good choice, if they are relatively cheap and non-invasive and if they reflect exposure over a relevant time period—usually months or years. An example is the use of blood lead measurements in a cohort study of occupational exposure to lead among lead smelter workers (Gerhardsson *et al.*, 1995).

For biomarkers that are more expensive to measure, an alternative is to use a nested case–control design. Here, biological samples are collected from all cohort members at baseline and stored. An important issue in investigations of this kind, therefore, is the stability of the markers during storage (Riboli *et al.*, 1995). Collected samples are

analysed only for those subjects who go on to develop a cancer of interest and for suitably chosen controls. This approach was used successfully in Shanghai in China (Ross *et al.*, 1992; Qian *et al.*, 1994) where a nested case–control analysis showed highly significant associations between the presence of urinary aflatoxins, serum hepatitis B surface antigen positivity and risk for hepatocellular carcinoma (HCC). A cohort analysis including the same cases of HCC revealed no strong or statistically significant association between HCC and dietary aflatoxin consumption, as determined using food frequency questionnaires and aflatoxin levels in local foods.

Case–control studies relating exposure to cancer

In this case the biomarker should reflect exposure over many years and must not be altered as a consequence of the disease process or its therapy. One such example is the measurement in serum of DDE as a marker for DDT exposure in a case–control study of breast cancer in the United States of America (USA) (Wolff *et al.*, 1993). This approach is feasible because organochlorine compounds such as DDT or PCBs accumulate in adipose tissue with an elimination half-life of several decades.

However, a biomarker with a shorter half-life may still be applicable in case–control studies, provided patterns of exposure remain fairly constant over time and are not modified by the development of disease. The potential for such use would be greatest in studies involving early, asymptomatic cases of cancer (Rothman *et al.*, 1995).

Ecological studies relating exposure to cancer

In ecological studies, exposures and disease outcomes are compared in populations or groups of people. For example, levels of *N*-nitrosamino acids in the overnight urine of men living in 69 counties

Table 2. Parameters proposed for biomonitoring exposure to carcinogens in occupationally exposed workers in four countries (IPCS, 1993)

IARC carcinogenicity evaluation	ACGIH, 1992	DFG, 1992	FIOH, 1993	UK HSE, 1991
Group 1				
Benzene	U-phenol	U-phenol, B-benzene	B-benzene	B-benzene
Ethylene oxide		B-ethylene oxide		
Vinyl chloride		U-thiodiglycolic acid		
Arsenic		U-volatile As compounds	U-As^{3+}, As^{5+}	U-As^{3+}, As^{5+}, MMA + DMA
Cadmium	U-Cd, B-Cd	U-Cd, B-Cd	U-Cd, B-Cd	U-Cd, B-Cd
Chromium	U-Cr	U-Cr	U-Cr	B-Cr, U-Cr
Nickel		U-Ni	U-Ni	U-Ni
Group 2a				
MOCA				U-MOCA
Tetratrachloroethylene	B-tetrachloroethylene	B-tetrachloroethylene		B-tetrachloroethylene
Trichloroethylene	U-TCA, B-trichloroethanol	U-TCA, B-trichloroethanol	U-TCA, U-trichloroethanol	U-TCA
Group 2b				
Cobalt		U-Co	U-Co	U-Co
Lead	B-Pb, U-Pb, B-ZPP	B-Pb, U-ALA	B-Pb, B-ZPP	B-Pb, U-ALA, B-ZPP

Abbreviations: ACGIH, American Conference of Governmental Industrial Hygienists; DFG, Deutsche Forschungsgemeinschaft; FIOH, Finnish Institute of Occupational Health; UK HSE, United Kingdom Health and Safety Executive; U, urine; B, blood; MMA, monomethylarsinic acid; DMA, dimethylarsinic acid; MOCA, methylene *bis*(2-chloroaniline); TCA, trichloroacetic acid; ALA, δ-aminolevulinic acid; ZPP, erythrocyte zinc protoporphyrin.

in China have been geographically associated with oesophageal cancer mortality (Wu *et al.*, 1993). For this type of study, some imprecision in the measurement of biomarkers can be tolerated, provided that a relatively large number of samples are measured in each geographical unit, since analysis is based on estimates of population means.

Cross-sectional studies linking exposure with early markers of carcinogenesis

Cross-sectional studies can be used to examine whether suspected carcinogenic exposures are associated with early markers of carcinogenesis, such as chromosomal aberrations. Biomarkers of internal dose sometimes provide a convenient measure of exposure for such investigations. For example, a number of studies looking at the association of styrene with cytogenetic damage in lymphocytes have used urinary mandelic acid as a marker for exposure (IARC, 1994). For a study designed in this way, the biomarker of exposure should also cover a period relevant to the early marker of carcinogenesis (months or years).

Studies to assess the contribution of different sources of exposure to total dose

Epidemiology may be used to investigate not only the risk of cancer from different levels of exposure to a carcinogen, but also the importance of different sources of exposure in determining total dose. In this type of study, biomarkers of internal dose which relate only to recent exposure are often the most useful. For example, concerns have been expressed that clusters of leukaemia in the vicinity of petrochemical works might be attributable to releases of benzene into ambient air. However, studies by Wallace *et al.* (1987) in the USA indicate that residence near petrochemical works has negligible impact on personal doses of benzene (assessed by measurement in exhaled breath). Far more important is whether the individual smokes or is exposed to environmental tobacco smoke and how he/she makes use of motor vehicles.

Studies to evaluate other methods of exposure measurement

Even where a biomarker cannot be used on a large scale (e.g. because of expense or ethical constraints), it may still be useful in the evaluation or refinement of other methods of exposure assessment. For example, in a retrospective study of people manufacturing phenoxy herbicides and chlorophenols (Manz *et al.*, 1991; Flesch Janys *et al.*, 1995), exposure to dioxins was assessed from job histories and information about processes and operating methods at the plant where they worked. Blood and adipose tissue levels of dioxins, measured in a subset of the cohort, confirmed the departments with the highest exposure and allowed quantitative estimates of personal exposure for all members of the cohort.

Studies have also been carried out to compare exposure levels measured by different analytical approaches. For example, urinary levels of 1-hydroxypyrene, a biomarker for exposure to PAH, were shown to correlate well with levels of PAH–DNA adducts in the white blood cells of aluminium workers (van Schooten *et al.*, 1995), but not in foundry workers exposed to lower levels of PAH (Santella *et al.*, 1993) .

Conclusion

In summary, biomarkers of internal dose have found successful application in cancer epidemiology. The choice of a biomarker of internal dose for a particular epidemiological study or type of study requires careful consideration of the period of exposure to which the biomarker relates, host factors related to carcinogen metabolism, invasiveness of sampling, reliability and the cost of the assay. Before a new biomarker is adopted, these characteristics must be assessed in transitional studies to ensure that it will be applied appropriately.

References

Armstrong, B.K., White, E. & Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology*, Oxford, Oxford University Press

De Waard, F., Kemmeren, J.M., van Ginkel, L.A. & Stolker, A.A. (1995) Urinary cotinine and lung cancer risk in a female cohort. *Br. J. Cancer*, 72, 784–787

Enterline, P.E. & Marsh, G.M. (1982) Cancer among workers exposed to arsenic and other substances in a copper smelter. *Am. J. Epidemiol.*, 116, 895–911

Flesch Janys, D., Berger, J., Gurn, P., Manz, A., Nagel, S., Waltsgott, H. & Dwyer, J.H. (1995) Exposure to polychlorinated dioxins and furans (PCDD/F) and mortality in a cohort of workers from a herbicide-producing plant in Hamburg, Federal Republic of Germany. *Am. J. Epidemiol.*, 142, 1165–1175

Gerhardsson, L., Hagmar, L., Rylander, L. & Skerfving, S. (1995) Mortality and cancer incidence among secondary lead smelter workers. *Occup. Environ. Med.*, 52, 667–672

Ghezzi, I., Toffoletto, F., Sesana, G., Fagioli, M.G., Micheli, A., Di Silvestro, P., Zocchetti, C. & Alessio, L. (1985) Behaviour of biological indicators of cadmium in relation to occupational exposure. *Int. Arch. Occup. Environ. Health*, 55, 133–140

Groopman, J.D. & Skipper, P.L. (1991) *Molecular Dosimetry and Human Cancer: Analytical, Epidemiological, and Social Considerations*, Boca Raton, CRC Press

Huang, M.Y., Jin, C., Liu, Y.T., Li, B.H., Qu, Q.S., Uchida, Y., Inoue, O., Nakatsuka, H., Watanabe, T. & Ikeda, M. (1994) Exposure of workers to a mixture of toluene and xylenes. I. Metabolism. *Occup. Environ. Med.*, 51, 42–46

Hulka, B.S. (1991) Epidemiological studies using biological markers: issues for epidemiologists. *Cancer Epidemiol. Biomarkers. Prev.*, 1, 13–19

Hulka, B.S., Wilcosky, T.C. & Griffith, J.D. (1990) *Biological Markers in Epidemiology*, Oxford, Oxford University Press

IARC (1994) *IARC Monographs on the Evaluation of Carcinogenic Risk to Humans, Some Industrial Chemicals*, Vol. 60, pp. 233–320

Inoue, O., Seiji, K., Kawai, T., Watanabe, T., Jin, C., Cai, S.X., Chen, Z., Qu, Q.S., Zhang, T. & Ikeda, M. (1993) Excretion of methylhippuric acids in urine of workers exposed to a xylene mixture: comparison among three xylene isomers and toluene. *Int. Arch. Occup. Environ. Health*, 64, 533–539

IPCS, (1993) *Biomarkers and Risk Assessment: Concepts and Principles.* (Environmental Health Criteria 155), Geneva, World Health Organization

Kang, D.H., Rothman, N., Cho, S.H., Poirier, M.C., Greenberg, A., Hsu, C.H., Schwartz, B.S., Baser, M.E., Weston, A., Groopman, J.D. & Strickland, P.T. (1995) Interindividual differences in the concentration of 1-hydroxypyrene-glucuronide in urine and polycyclic aromatic hydrocarbon-DNA adducts in peripheral white blood cells after charbroiled beef consumption. *Carcinogenesis*, 16, 1079–1085

Krieger, N., Wolff, M.S., Hiatt, R.A., Rivera, M., Vogelman, J. & Orentreich, N. (1994) Breast cancer and serum organochlorines: a prospective study among white, black, and Asian women. *J. Natl Cancer Inst.*, 86, 589–599

Manz, A., Berger, J., Dwyer, J.H., Flesch Janys, D., Nagel, S. & Waltsgott, H. (1991) Cancer mortality among workers in chemical plant contaminated with dioxin. *Lancet*, 338, 959–964

Qian, G.S., Ross, R.K., Yu, M.C., Yuan, J.M., Gao, Y.T., Henderson, B.E., Wogan, G.N. & Groopman, J.D. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol. Biomarkers Prev.*, 3, 3–10

Riboli, E., Haley, N.J., De Waard, F. & Saracci, R. (1995) Validity of urinary biomarkers of exposure to tobacco smoke following prolonged storage. *Int. J. Epidemiol.*, 24, 354–358

Ross, R.K., Yuan, J.M., Yu, M.C., Wogan, G.N., Qian, G.S., Tu, J.T., Groopman, J.D., Gao, Y.T. & Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, 339, 943–946

Rothman, N., Stewart, W.F. & Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev.*, 4, 301–311

Santella, R.M., Hemminki, K., Tang, D.L., Paik, M., Ottman, R., Young, T.L., Savela, K., Vodickova, L., Dickey, C., Whyatt, R., *et al.* (1993) Polycyclic aromatic hydrocarbon-DNA adducts in white blood cells and urinary 1-hydroxypyrene in foundry workers. *Cancer Epidemiol. Biomarkers Prev.*, 2, 59–62

Schulte, P.A. & Perera, F.P. (1993) *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press

Sherson, D., Sigsgaard, T., Overgaard, E., Loft, S., Poulsen, H.E. & Jongeneelen, F.J. (1992) Interaction of smoking, uptake of polycyclic aromatic hydrocarbons, and cytochrome P450IA2 activity among foundry workers. *Br. J. Ind. Med.*, 49, 197–202

Sinha, R., Rothman, N., Brown, E.D., Mark, S.D., Hoover, R.N., Caporaso, N.E., Levander, O.A., Knize, M.G., Lang, N.P. & Kadlubar, F.F. (1994) Pan-fried meat containing high levels of heterocyclic aromatic amines but low levels of polycyclic aromatic hydrocarbons induces cytochrome P4501A2 activity in humans. *Cancer Res.*, 54, 6154–6159

van Schooten, F.J., Jongeneelen, F.J., Hillebrand, M.J., van Leeuwen, F.E., de Looff, A.J., Dijkmans, A.P., van Rooij, J.G., Den Engelse, L. & Kriek, E. (1995) Polycyclic aromatic hydrocarbon-DNA adducts in white blood cell DNA and 1-hydroxypyrene in the urine from aluminum workers: relation with job category and synergistic effect of smoking. *Cancer Epidemiol. Biomarkers Prev.*, 4, 69–77

Wallace, L.A., Pellizzari, E.D., Hartwell, T.D., Sparacino, C., Whitmore, R., Sheldon, L., Zelon, H. & Perritt, R. (1987) The TEAM (Total Exposure Assessment Methodology) Study: personal exposures to toxic substances in air, drinking water, and breath of 400 residents of New Jersey, North Carolina, and North Dakota. *Environ. Res.*, 43, 290–307

Wolff, M.S., Toniolo, P.G., Lee, E.W., Rivera, M. & Dubin, N. (1993) Blood levels of organochlorine residues and risk of breast cancer. *J. Natl Cancer Inst.*, 85, 648–652

Wu, Y., Chen, J., Ohshima, H., Pignatelli, B., Boreham, J., Li, J., Campbell, T.C., Peto, R. & Bartsch, H. (1993) Geographic association between urinary excretion of *N*-nitroso compounds and oesophageal cancer mortality in China. *Int. J. Cancer*, 54, 713–719

Corresponding author

D. Coggon
MRC Environmental Epidemiology Unit,
University of Southampton,
Southampton, UK

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Biochemical markers of dietary intake

R. Kaaks, E. Riboli and R. Sinha

The primary objective of nutritional epidemiology is to identify, in combination with other forms of research, which aspects of diet and nutritional factors are causally related to cancer development. However, traditional epidemiology can evaluate with only a limited degree of specificity to which individual dietary factors an increased occurrence of cancer can be attributed. The two main reasons for this are: (1) dietary intake levels of specific foods or food constituents can be strongly intercorrelated; (2) dietary intake levels of specific food constituents are generally measured with rather large errors. Biochemical markers are increasingly seen as measurements that may help to overcome some of the above-mentioned methodological problems in nutritional epidemiology.

The most promising application of biomarkers of diet is in prospective cohort studies, where the marker is unlikely to have been influenced by the disease, as biological specimens can be collected and stored well before cancer develops and becomes clinically manifest.

Two main types of markers of diet should be distinguished:

1. Markers of absolute quantitative intake levels such as urinary nitrogen for protein. These markers are based on a time-related balance between intake and output.
2. Markers based on the concentration of a specific substance in biological fluids, or tissues (e.g. the concentration of vitamins in blood plasma). Such markers have no time dimension, and provide only a correlate of dietary intake level. This class represents the vast majority of biomarkers of diet.

The main advantages of biomarkers of diet are as follows:

- Biomarkers may provide more accurate measurements—or at least a correlate of intake levels of specific chemical constituents—than the traditional dietary assessment methods.
- Biomarkers are objective measurements, the validity and precision of which are independent of the subjects' memory or capacity to describe foods consumed. In statistical terms, this means that the measurement errors of biomarkers can be assumed to be uncorrelated with those of the questionnaire measurements. It is this statistical independence of these errors that makes biomarkers of diet such an attractive class of measurement in epidemiological studies.

In prospective cohort studies, biomarkers of diet can be used as an independent measurement of dietary intakes in order to evaluate the association between these intakes and disease risk, or as an additional measurement of diet to be combined statistically (e.g. in latent variable models) with measurements obtained by questionnaire. This latter application is especially useful in validity studies to estimate the magnitude of systematic and random errors of dietary exposure measurements obtained from questionnaires.

The primary objective of nutritional epidemiology is to identify, in interaction with other forms of research, to which specific aspects of diet or other nutritional lifestyle factors (e.g. physical activity) observed variations in incidence of disease can be attributed. In theory, by identifying causal factors with a highest possible specificity, accurate recommendations can be given about nutrition-related lifestyle factors, with the aim of reducing the overall incidence of disease. Besides providing insight into how to optimize nutritional lifestyle patterns, it is hoped that epidemiological research may also help to identify individual chemical agents that, even when administered as isolated compounds, may reduce cancer risk. Examples of agents for which such potentially 'chemopreventive' action has been postulated include vitamins C, E and A, carotenoids, flavonoids, indoles or phytoestrogens (Steinmetz

& Potter, 1991; Wattenberg, 1992). In practice, however, epidemiological studies can evaluate with only a limited degree of specificity to which individual dietary risk factors (foods, nutrients or other chemical food constituents) an increased occurrence of disease can be attributed. The two major reasons for this are: (1) intake levels of specific foods, food groups or chemical food constituents (nutrients or non-nutrient substances) can be strongly intercorrelated; and (2) the intake levels of specific foods or food constituents are generally measured with rather large errors.

In nutritional epidemiology, biochemical markers are increasingly seen as measurements that may help to overcome some of the above-mentioned methodological problems in nutritional epidemiology. In this chapter, we discuss the potential uses and limitations of biochemical markers of dietary intake levels in studies of the relationships between diet and the risk of chronic diseases such as cancer. Although much of this discussion will apply to the use of biomarkers in practically any type of epidemiological design, including ecological correlation studies and case–control studies, we shall focus mainly on their use in prospective cohort studies. Prospective cohort studies have the advantage that the level of the marker is unlikely to have been influenced by the presence or absence of disease, as biological specimens can be collected well before disease develops or becomes clinically manifest. A new generation of prospective cohort studies on diet and cancer, including large banks of biological specimens (blood, urine, toenails), is currently being conducted in different parts of the world (Toniolo *et al.*, 1995; Berrino *et al.*, 1996; Riboli & Kaaks, 1997).

Types of biomarker of diet

Most markers currently in use are related to the intake of a specific chemical compound ingested together with a food (Riboli *et al.*, 1987). These compounds may be natural food constituents (e.g. the traditional nutrients); they may be formed during processing or treatment (e.g. heterocyclic amines formed in foods: PhIP, MeIQ; transfatty acids); or they may be contaminants from either a natural origin (e.g. aflatoxin) or a xenobiotic origin (e.g. DDT, PCBs). From a methodological point of view, it is useful to distinguish two major classes of biomarkers of diet.

The first class are markers based on knowledge about the metabolic balance between the intake and excretion of specific chemical components; that is, the per cent recovery of the compound or its metabolites in excretion products (mainly urine and breath) is known. All markers belonging to this first category are time-related; that is, the markers are based on a balance between intake and output over a representative time period, usually of 24 h, and thus can be translated into estimates of absolute intake level over 24 h as well. Probably the best known example is the 24-h urinary excretion of nitrogen as a marker of average 24-h protein intake (Bingham & Cummings, 1985). This marker is based on the knowledge that, in individuals who are in nitrogen balance (i.e., in practical terms, individuals who have a stable body mass and composition), the 24-h urinary nitrogen excretion represents a practically constant proportion of nitrogen ingested (mostly in the form of protein). Likewise, the urinary excretion of potassium can be used as an indicator of potassium intake, even though the percentage recovery in urine of ingested potassium is more variable between individuals than the recovery of nitrogen (Bingham *et al.*, 1992, 1995). A high intake of potassium can be taken as an indicator of a diet rich in vegetables and fruits (Williams & Bingham, 1986). A third, well-known example of a biomarker based on a metabolic balance between ingestion and excretion is the 'doubly labelled water' method to estimate average daily energy expenditure (Schoeller, 1988). If, on average, a group of individuals is in energy balance (i.e. there are no gains or losses in body weight or composition), on average the daily energy expenditure is equal to daily energy intake; thus, the doubly labelled water method can be used to estimate dietary energy intake.

The second class of markers of dietary intake are measured as concentrations of specific substances in: biological fluids (plasma, urine, saliva), specific tissues or cells (e.g. adipose tissue, white blood cells), lipoproteins, cellular membranes, DNA and specific proteins. Here, the word 'concentration' must be taken in a broad sense, in that the denominator determining the concentration does not need to be measured in units of volume. Examples of different types of 'concentration' are the fraction (ppm) of DNA bases with specific adducts (e.g. of PhIP), the fraction of amino acids in a given

type of protein (albumin, haemoglobin) that carries a specific type of adduct (PhIP, aflatoxin), the concentration of vitamin E in lipoproteins or in the lipid phase of cell membranes, and the fatty acid composition of an adipose tissue biopsy. This second class of marker represents the vast majority of markers of dietary intake and, contrary to the markers based on a balance over time between intake and excretion, are without a time dimension. Also in contrast with the first class of markers, the quantitative relationship between this second class of markers and dietary intake level cannot be as clearly defined from knowledge about its physiology, and this quantitative relationship may vary between populations, or even between population subgroups, depending on the presence and relative impact of determinants other than diet. The markers belonging to this second category can therefore only provide a correlate of dietary intake level; that is, they can provide a more or less accurate relative ranking of individuals by intake level and/or disease risk, but cannot be translated directly into absolute (recommended) intake levels. A list of examples of nutritional markers belonging to this second category that are often used in epidemiology is given in Table 1.

Advantages of biomarkers as measurements of dietary intake

To appreciate the potential advantages of biochemical markers as measurements of diet, it is useful to review briefly the characteristics and limitations of the more traditional measurements of dietary intake level. The latter can be divided into assessments of an individual's *habitual*, long-term intake of foods, by means of a structured questionnaire or interview (Cameron & van Staveren, 1988; Willett, 1990a), and measurements of the *actual* consumption of foods on one or more specific days by means of written food consumption records (Cameron & van Staveren, 1988) or by means of short-term (usually 24-h) recalls (Witschi, 1990).

Food composition tables are used to translate either type of measurement of food consumption into estimates of nutrient intake. In case–control or cohort studies on the relationship between diet and the risk of developing chronic disease, measurements of habitual exposure levels can usually be obtained only by the first type of measurement (i.e. using a questionnaire); nevertheless, replicate measurements of the second type (written records or short-term recalls) can be used in smaller substudies for the 'validation' and/or 'calibration' of the questionnaire measurements. Either type of measurement—whether based on a questionnaire about habitual food consumption or on the recording or recall of the actual consumption on given days—can be biased as a result of unknown, or unquantifiable, subject-associated factors such as a lack of memory, an incapacity to identify or describe accurately types of food consumed, or a tendency to deviate from the habitual diet on days that written records of food consumption are being kept. When measurements are repeated by the same method, individuals may tend to make errors of a similar size and magnitude on two or more occasions. Therefore, errors may be correlated positively, i.e. the measurements may have less than perfect validity (see below). Between different types of dietary assessment method—e.g. between questionnaires of habitual dietary intake and records or short-term recalls of actual dietary intake levels on specific days—errors are often assumed to be practically independent. This assumption will be questionable, however, whenever some of the potential sources of error are the same. For example, subjects may vary in their motivation to provide complete answers to each type of method, and may thus underestimate dietary intake levels to a similar degree by both approaches (i.e. again resulting in a positive correlation of errors between the two types of measurement).

An initial reason for using biochemical markers of dietary intake level is that they may provide a more accurate measurement (or at least a correlate) of intake levels of specific chemical constituents than can be obtained by the traditional methods. The inaccuracy of the traditional methods may be due, for instance, to the lack of reliable food composition data. Food composition tables generally contain figures only of those chemical food constituents ('essential nutrients') that provide energy or that must be consumed regularly to prevent specific, known deficiency diseases that would otherwise develop relatively quickly. Although in many countries databases on food composition are currently being extended so as to include a wider range of chemical compounds (i.e. extending

Table 1 Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Vitamin A and carotenoids					
Roidt *et al.*, 1988	Total preformed retinol	302 men and women	All	Serum retinol	0.08
				Serum α-carotene	0.00
				Serum β-carotene	0.03
	Total vitamin A	302 men and women	All	Serum retinol	0.08
	Total dietary carotenoids	302 men and women		Serum retinol	0.02
				Serum α-carotene	0.26
				Serum β-carotene	0.21
Romieu *et al.*, 1990	Carotenoids	370 men and women	Non-smokers	Plasma β-carotene	0.37
van't Veer *et al.*, 1990, 1993	Fruits and vegetables	140 women	All (dietary history)	Plasma carotenoids	0.22
Coates *et al.*, 1991	Total carotenoids	50 women	Non-smokers	Serum provitamin A carotenoids	0.45
	Total carotenoids			Serum β-carotene	0.43
	Provitamin A carotenoids			Serum provitamin A carotenoids	0.38
	α-Carotene			Serum α-carotene	0.38
	β-Carotene			Serum β-carotene	0.32
	Cryptoxanthine			Serum cryptoxanthine	0.36
	Lycopene			Serum lycopene	−0.06
	Lutein			Serum lutein/zeaxanthin	0.00
	Total carotenoids	41 women	Smokers	Serum provitamin A carotenoids	0.17
	Total carotenoids			Serum β-carotene	0.23
	Provitamin A carotenoids			Serum provitamin A carotenoids	0.17
	α-Carotene			Serum α-carotene	0.21
	β-Carotene			Serum β-carotene	0.23
	Cryptoxanthine			Serum cryptoxanthine	0.18
	Lycopene			Serum lycopene	0.00
	Lutein			Serum lutein/zeaxanthin	0.08

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Ascherio *et al.*, 1992	Carotenoids	110 men	Non-smokers	Plasma α-carotene	0.52
				Plasma β-carotene	0.34
				Plasma α- plus β-carotene	0.34
				Plasma lycopene	0.13
				Plasma lutein	0.36
				Plasma zeaxanthin	0.11
				Plasma retinol	0.11
		162 women	Non-smokers	Plasma α-carotene	0.37
				Plasma β-carotene	0.30
				Plasma α- plus β-carotene	0.30
				Plasma lycopene	0.01
				Plasma lutein	0.19
				Plasma zeaxanthin	0.02
				Plasma retinol	0.13
Jacques *et al.*, 1993	Vitamin A	42 men and 71 women	Non-supplement users	Plasma retinol	0.02
		55 men and 82 women	Supplement and non-supplement users	Plasma retinol	0.07
Jacques *et al.*, 1993	Carotenoids	53 men and 81 women	Non-supplement users	Plasma carotenoids	0.29
			Supplement and non-supplement users	Plasma carotenoids	0.37
Campbell *et al.*, 1994	Total fruits and vegetables	50 men and 49 women	All (FFQ)	Plasma lutein	0.39
				Plasma β-carotene	0.44
				Plasma lycopene	−0.04
				Plasma α-carotene	0.54
				Plasma β-carotene	0.43
				Plasma sum of carotenoids	0.54
Hebert *et al.*, 1994	Retinol	167 men and women	All	Plasma retinol	0.14
				Plasma β-carotene	−0.02
	Total β-carotene			Plasma β-carotene	−0.17
				Plasma retinol	0.29

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Yong *et al.*, 1994	α-Carotene	98 women	All (7-day records)	Plasma α-carotene	0.59
	β-Carotene			Plasma β-carotene	0.15
	β-Cryptoxanthin			Plasma β-cryptoxanthin	0.49
	Lutein			Plasma lutein	0.29
	Lycopene			Plasma lucopene	0.41
	Total			Plasma total	0.51
	α-Carotene		All (FFQ)	Plasma α-carotene	0.52
	β-Carotene			Plasma β-carotene	0.44
	β-Cryptoxanthin			Plasma β-cryptoxanthin	0.30
	Lutein			Plasma lutein	0.29
	Lycopene			Plasma lutein	0.28
	Total			Plasma total	0.43
Porrini *et al.*, 1995	β-Carotene	11 men and 33 women	All (FFQ)	Plasma β-carotene	−0.07
		All (7-day records)		Plasma β-carotene	0.44
Enger *et al.*, 1995	Carotenoids	215 men and women	All	Plasma α-carotene	0.27
				Plasma β-carotene	0.22
				Plasma β-cryptoxanthin	0.35
				Plasma lutein + zeaxanthin	0.33
				Lycopene	0.36
Kardinaal *et al.*, 1995	Retinol	82 men and women	All (FFQ)	Plasma retinol	0.06
	β-Carotene			Plasma β-carotene	0.15
Vitamin E					
Roidt *et al.*, 1988	Vitamin E	50 women	Non-smokers—food source only	Serum α-tocopherol	0.36
			Non-smoker—total	Serum α-tocopherol	0.38
		41 women	Smokers—food source only	Serum α-tocopherol	0.02
			Smokers—total	Serum α-tocopherol	0.32
Romieu *et al.*, 1990	Vitamin E	339 men and women	Non-supplement users	Plasma α-tocopherol	0.16

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Sinha *et al.*, 1993	Vitamin E	65 men	All	Plasma α-tocopherol	0.32
		46 men	Non-supplement user	Plasma α-tocopherol	0.17
		19 men	Supplement user	Plasma α-tocopherol	0.37
		65 men	All	Plasma γ-tocopherol	–0.33
		46 men	Non-supplement user	Plasma γ-tocopherol	–0.08
		19 men	Supplement user	Plasma γ-tocopherol	–0.25
Jacques *et al.*, 1993	Vitamin E	41 men and 70 women	Non-supplement users	Plasma α-tocopherol	0.35
		55 men and 70 women	Supplement and non-supplement users	Plasma α-tocopherol	0.53
Porrini *et al.*, 1995	Total α-tocopherol	11 men and 33 women	All (measured by FFQ)	Plasma α-tocopherol	–0.22
			All (measured by 7-day record)	Plasma α-tocopherol	0.10
Kardinaal *et al.*, 1995	α-Tocopherol	82 men and women	All (FFQ)	Plasma α-tocopherol	0.11
	α-Tocopherol: cholesterol ratio			Plasma α-tocopherol: cholesterol ratio	0.22
Other vitamins					
Jacques *et al.*, 1993	Vitamin D	42 men and 66 women	Non-supplement users	Plasma 25-OH vitamin D	0.25
		55 men and 80 women	Supplement and non-supplement users	Plasma 25-OH vitamin D	0.35
	Thiamin	42 men and 67 women	Non-supplement users	Red blood cell thiamin	0.01
		56 men and 82 women	Supplement and non-supplement users	Red blood cell thiamin	0.02
	Vitamin B_2	42 men and 66 women	Non-supplement users	Red blood cell riboflavin	–0.21
			Supplement and non-supplement users	Red blood cell riboflavin	–0.13
	Vitamin B_6 (pyridoxine)	42 men and 66 women	Non-supplement users	Red blood cell B_6	0.05

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Jacques *et al.* 1993 (Contd)	Vitamin B_6 (pyridoxine)	56 men and 82 women	Supplement and non-supplement users	Red blood cell B_6	−0.15
	Vitamin B_{12} (cyanocobalamine)	40 men and 65 women	Non-supplement users	Plasma B_{12}	0.19
		54 men and 78 women	Supplement and non-supplement users	Plasma B_{12}	0.35
	Plasma folate	40 men and 65 women	Non-supplement users	Plasma folate	0.61
			Supplement and non-supplement users	Plasma folate	0.63
Wild *et al.*, 1994	Folate	16 women	Control subjects	Serum folate	0.81
		16 women	Women with infants with neural tube defect	Serum folate	−0.29
		16 women	Control subjects	Red blood cell folate	0.72
			Women with infants with neural tube defect	Red blood cell folate	−0.08
Sinha *et al.*, 1994	Vitamin C	493 women	From food	Serum ascorbic acid	0.19
			From food + supplement	Serum ascorbic acid	0.32
			From food (one 24-h recall)	Serum ascorbic acid	0.36
			From food + supplement	Serum ascorbic acid	0.56
Sinha *et al.*, 1992	Vitamin C	68 men	All	Plasma ascorbic acid	0.43
Jacques *et al.*, 1993	Vitamin C	39 men and 60 women	Non-supplement users	Plasma total vitamin C	0.38
		55 men and 80 women	Supplement and non-supplement users	Plasma total vitamin C	0.43
Porrini *et al.*, 1995	Ascorbic acid	11 men and 33 women	From food (FFQ)	Whole blood ascorbic acid	−0.22
			From food (7-day records)	Whole blood ascorbic acid	0.44

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Protein					
O'Donnell *et al.*, 1991	Protein	24 men	All (16-day weighed records)	Serum urea	0.21
				Serum creatinine	0.12
				Urine nitrogen (24-h)	0.18
			All (FFQ)	Serum urea	−0.004
				Serum creatinine	−0.07
				Urine nitrogen (24-h)	0.36
		28 women	All (16-day weighed records)	Serum urea	0.07
				Serum creatinine	0.39
				Urine nitrogen (24-h)	0.19
			All (FFQ)	Serum urea	0.09
				Serum creatinine	0.12
				Urine nitrogen (24-h)	0.28
Selenium					
Snook *et al.*, 1983	Selenium	155 men and women	All (three 24-hr recalls)	Plasma selenium	0.12
Yang *et al.*, 1989	Selenium	Approx. 150 families	3-day duplicate plate measured for selenium	Whole blood selenium	0.88
				Breast milk	0.90
				24-hr urine	0.86
Swanson *et al.*, 1990	Selenium	24 men and 20 women	Minimum of six duplicate plates for selenium measurement	Serum selenium	0.51
				Whole blood selenium	0.55
				Toenail selenium	0.53
				Urine selenium	0.87
van't Veer *et al.*, 1990, 1993	Selenium	243 women	All (dietary history)	Plasma selenium	0.15
		239 women		Erythrocyte selenium	−0.05
		360 women		Toenail selenium	−0.01
Other minerals					
O'Donnell *et al.*, 1991	Sodium	24 men	All (16-day weighed records)	Urine sodium	0.008
	Potassium			Urine potassium	0.19
	Zinc			Serum zinc	-0.08
	Sodium		All (FFQ)	Urine sodium	0.25
	Potassium			Urine potassium	0.08
	Zinc			Serum zinc	0.16

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
O'Donnell *et al.* 1991 (Contd)	Sodium	28 women	All (16-day weighed records)	Urine sodium	0.09
	Potassium			Urine potassium	0.44
	Zinc			Serum zinc	–0.02
	Sodium		All (FFQ)	Urine sodium	0.24
	Potassium			Urine potassium	0.04
	Zinc			Serum zinc	0.32
Jacques *et al.*, 1993	Magnesium	39 men and 70 women	Non-supplement users	Serum magnesium	0.15
		56 men and 82 women	Supplement and non-supplement users	Serum magnesium	0.27
	Zinc	44 men and 76 women	Non-supplement users	Serum zinc	0.10
		56 men and 82 women	Supplement and non-supplement users	Serum zinc	0.11
Matkovic *et al.*, 1995	Calcium	381 pre-teen girls	All (3-day diary)	Urine calcium	0.21
	Sodium			Urine sodium	0.21
Donovan & Gibson, 1995	Iron	124 women	All (3-day weighed records)	Serum iron	0.26
	Zinc			Serum zinc	0.16
Shaw *et al.*, 1995	Iron	43 men and 71 women	All (6-day weighed food records)	Plasma haemoglobin	0.26
Fatty acids					
van Staveren *et al.*, 1986	P:S ratio[a]	162 men and women	All subjects (2-day records)	Adipose P:S ratio	0.38
	M:P ratio			Adipose M:P ratio	0.38
	P			Adipose P	0.40
	P:S ratio	59 women	Average 19 24-h recalls	Adipose P:S ratio	0.57
	M:P ratio			Adipose M:P ratio	0.63
	P			Adipose P	0.68
	Linoleic:S ratio			Adipose linoleic: S ratio	0.62
	M:linoleic ratio			Adipose M:linoleic ratio	0.63
	Linoleic acid			Adipose linoleic	0.70

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
London *et al.*, 1991	S	115 postmenopausal women	FFQ	Adipose S	0.02
	M			Adipose M	0.08
	Oleic acid			Adipose oleic acid	0.06
	Trans fatty acids			Adipose *trans* fatty acids	0.41
	P			Adipose P	0.15
	Linoleic acid			Adipose linoleic acid	0.13
	Linolenic acid			Adipose linolenic acid	0.07
	n–3 fatty acids of marine origin			Adipose *n*–3 fatty acids of marine origin	0.43
	P/S ratio			Adipose P/S ratio	0.28
Hunter *et al.*, 1992	S	118 men	FFQ	Adipose S	0.14
	M			Adipose M	0.21
	P			Adipose P	0.10
	P:S ratio			Adipose P:S ratio	0.43
	Palmitic			Adipose palimitic	0.21
	Oleic			Adipose oleic	0.28
	Linoleic			Adipose linoleic	0.10
	Eicosapentaenoic			Adipose eicosapentaenoic	0.43
	Trans-isomers			Adipose *trans*-isomers	0.21
	S	118 men	2 weeks of diet records	Adipose S	0.25
	M			Adipose M	0.16
	P			Adipose P	0.09
	P:S ratio			Adipose P:S ratio	0.40
Feunekes *et al.*, 1993	P:S ratio[a]	55 men and women	FFQ	Adipose P:S ratio	0.32
	M:P ratio			Adipose M:P ratio	0.24
	P			Adipose P	0.24
	Linoleic:S ratio			Adipose linoleic: S ratio	0.33
	M:linoleic ratio			Adipose M:linoleic ratio	0.25
	Linoleic acid			Adipose linoleic	0.28
	P:S ratio[a]		Dietary history	Adipose P:S ratio	0.28
	M:P ratio			Adipose M:P ratio	0.34
	P			Adipose P	0.29
	Linoleic/S ratio			Adipose linoleic: S ratio	0.30

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Feunekes *et al.* 1993 (Contd)	M:linoleic ratio			Adipose M:linoleic ratio	0.44
	Linoleic acid			Adipose linoleic	0.34
	P:S ratio[a]	99 men and women	FFQ	Erythrocyte membrane P:S ratio	0.22
	M:P ratio			Erythrocyte membrane M:P ratio	0.37
	P			Erythrocyte membrane P	0.33
	Linoleic:S ratio			Erythrocyte membrane linoleic: S ratio	0.40
	M:linoleic ratio			Erythrocyte membrane M:linoleic ratio	0.41
	Linoleic acid			Erythrocyte membrane linoleic	0.44
Tjønneland *et al.*, 1993	S	67 men and 121 women	FFQ	Adipose S	0.24
	M			Adipose M	0.05
	P			Adipose P	0.44
	18:2 *n*–6			Adipose 18:2 *n*–6	0.44
	19:3 *n*–3			Adipose 18:2 *n*–3	0.12
	20:5 *n*–3			Adipose 20:5 *n*–3	0.47
	22:6 *n*–3			Adipose 22:6 *n*–3	0.41
	S	67 men and 121 women	Two 7-day diaries	Adipose S	0.46
	M			Adipose M	0.19
	P			Adipose P	0.57
	18:2 *n*–6			Adipose 18:2 *n*–6	0.51
	19:3 *n*–3			Adipose 18:2 *n*–3	0.36
	20:5 *n*–3			Adipose 20:5 *n*–3	0.44
	22:6 *n*–3			Adipose 22:6 *n*–3	0.55
Hebert *et al.* 1994	Fat	167 men and women		Plasma cholesterol	0.02
				Plasma HDL-cholesterol	–0.12
	Cholesterol			Plasma cholesterol	0.02
				Plasma HDL-cholesterol	–0.05
	S			Plasma cholesterol	<0.01
				Plasma HDL-cholesterol	–0.12

Table 1 (Contd) Selected studies of relationships between biochemical markers of dietary intake and their correlations with intake levels, measured by food frequency questionnaire (FFQ) (unless otherwise stated) or by 7-day record or 24-h diet recall

Author	Nutrients or food groups	No. of subjects	Category of subjects	Biochemical index	Correlation coefficient
Hebert *et al.* 1994 (Contd)	M			Plasma cholesterol	0.02
				Plasma HDL-cholesterol	–0.12
	P			Plasma cholesterol	0.01
				Plasma HDL-cholesterol	–0.09
Ma *et al.*, 1995		3570 men and women	FFQ	Plasma cholesterol esters:	
	S			S	0.23
	M			M	0.01
	P			P	0.31
	P:S ratio			P:S ratio	0.30
	16:0			16:0	0.19
	18:1 *n*–9			18:1 *n*–9	0.03
	18:2 *n*–6			18:2 *n*–6	0.28
	18:3 *n*–3			18:3 *n*–3	0.21
	20:5 *n*–3			20:5 *n*–3	0.23
	22:6 *n*–3			22:6 *n*–3	0.42
				Plasma phospholipids:	
	S			S	0.15
	M			M	0.05
	P			P	0.25
	P:S ratio			P:S ratio	0.24
	16:0			16:0	0.16
	18:1 *n*–9			18:1 *n*–9	0.08
	18:2 *n*–6			18:2 *n*–6	0.22
	18:3 *n*–3			18:3 *n*–3	0.15
	20:5 *n*–3			20:5 *n*–3	0.20
	22:6 *n*–3			22:6 *n*–3	0.42

[a]P, polyunsaturated fatty acids; M, monounsaturated fatty acids; S, saturated fatty acids.

beyond the traditional nutrients), an important problem is that chemical constituents may have rather variable concentrations within given types of food, due to natural variations between cultivars (e.g. indoles in cabbage; Preobrazhenskaya *et al.*, 1993; Guengerich, 1995), due to variations in the soil content of specific minerals and trace elements (e.g. selenium; Ullrey, 1981), or due to effects of storage, processing and cooking methods (e.g. formation of aromatic amines such as PhIP, MeIQ; Sinha *et al.*, 1995; Skog *et al.*, 1995). Information on these sources of variation (type of cultivar, soil, cooking methods) often cannot be assessed reliably using questionnaire methods, and it will therefore be impossible to apply correct food composition values by which the intake of these specific food constituents can be computed from the estimated food consumption levels.

A second possible advantage of biomarkers often mentioned is their *objectiveness*; that is, the validity and precision of the markers can reasonably be assumed to be independent of the subjects' memory or capacity to describe foods consumed. Only markers based on 24-h urinary excretion of a given compound may be influenced directly by factors related to subjects' motivation, as urine collections may be incomplete. However, the completeness of urine collections can be checked by measuring the recovery of a small amount of PABA (para-amino benzoic acid), ingested in the form of small tablets at regular intervals during the 24-h collection period (Bingham & Cummings, 1986). It must be noted that the term 'objectiveness' does not necessarily mean that the validity and precision are the same for cases with a given disease and for disease-free controls; this may be true only if biological specimens are collected well before any metabolic changes occur that may be a consequence rather than a cause of disease and that may have influenced the level of the marker (i.e. using a prospective study design). The more relevant interpretation of the word objectiveness is that random errors in the marker—i.e. any variations that are not correlated with the individuals' true, habitual dietary intake levels—can be assumed to be statistically independent of subjects' capacity or motivation to give an accurate response to traditional methods of measurement. Thus, the errors of biomarkers can be assumed to be uncorrelated with those of questionnaire measurements, short-term recalls or food consumption records. It is this statistical independence of the errors which in theory makes biomarkers of diet a very attractive additional class of dietary intake measurement for use in epidemiological studies.

The use of biomarkers of diet in prospective cohort studies

A first type of application of biomarkers in nutritional epidemiology is their use as an independent measurement of dietary intake levels, to evaluate the association between these intake levels and disease risk. There are numerous examples of this type of application, for instance in studying the relationships between cancer risk and plasma levels of carotenoids, vitamin C, selenium or the fatty acid composition of blood lipids (e.g. Stähelin *et al.*, 1991; London *et al.*, 1992; Garland *et al.*, 1993; van den Brandt *et al.*, 1993; Zheng *et al.*, 1993; Berg *et al.*, 1994; Gann *et al.*, 1994; Kabuto *et al.*, 1994). As mentioned above, an initial argument for this type of application is that the marker may classify individuals more accurately than traditional dietary assessment methods by habitual intake level of some food constituents. For instance, dietary intake levels of selenium may not be calculated very reliably from individuals' estimated levels of food consumption, because selenium concentrations can vary widely in individual foods. Likewise, intake levels of specific types of carotenoids (e.g. lycopene, zeoxanthine) or specific types of fatty acid may not be estimable because of a lack of food composition data. A further argument for using a biomarker as a measurement of dietary intake level is that a relationship between a marker and disease risk may be interpreted as *independent* confirmation of a relationship found between disease risk and dietary intake measurements obtained by more traditional techniques (e.g. dietary questionnaires) that have different sources of error (although some caution should be made against oversimplistic interpretations of biomarkers as a measurement uniquely of dietary intake level; see the section on 'Validity and confounding' below).

A second type of application of biomarkers of diet is to combine it, as an *additional* measurement of diet, with measurements obtained by questionnaires, recalls or food consumption records. None of the existing methods of measuring individuals' true, habitual (long-term) intake levels of specific

foods or nutrients provides error-free measurements. Food frequency questionnaires are the method used most often to measure individuals' dietary intake levels at baseline in prospective cohort studies. Systematic and random errors in the questionnaire measurements lead to bias in estimates of relative risk or other measures of diet–disease association. By estimating the average magnitudes of these errors, corrections may be made for the bias in relative risk estimates. For instance, assume that questionnaire measurements are related linearly to a true intake level T, i.e.

$$Q = \alpha_Q + \beta_Q T + \varepsilon_Q$$

Unknown parameters in this model are the constant and proportional scaling factors, α_Q and β_Q, the mean (μ_T) and variance (σ^2_T) of the population distribution of true intake values T, and the population variance $\sigma^2_{\varepsilon Q}$ of random measurement errors ε_Q. From estimates of the parameters β_Q, $\sigma^2_{\varepsilon Q}$ and σ^2_T, the coefficient of correlation ρ_{QT} between questionnaire measurements and true intake levels can be computed. When relative risks are calculated for quantile levels (e.g. quintiles) of the population distribution of intake levels, this correlation coefficient determines the degree of underestimation ('attenuation bias') due to random errors in the questionnaire measurements, and the loss in statistical power that such underestimation entails (de Klerk *et al.*, 1989).

By making comparisons between questionnaire measurements and other types of dietary intake assessment in *validity substudies*, it may be possible to estimate key parameters α_Q, β_Q, μ_T, σ^2_T, and $\sigma^2_{\varepsilon Q}$ in the measurement model above. Estimation of these parameters then relies on statistical models in which true intake levels, T, are considered to be values of a *latent* variable. A crucial assumption for the parameters to be estimated is that any association between complementary measurements of diet must be due *only* to the relationship between each of these variables and the same latent variable (true dietary intake level); that is, variations that are uncorrelated with true intake level ('random errors') should also be uncorrelated between the different types of measurement. Generally, such independence of errors cannot be proven in practice, but only assumed (although, in more complex analyses of reliability studies with replicate measurements of different types, assuming independence of errors between some, minimum, number of variables may sometimes allow the estimation of correlations of some other errors). To increase the likelihood that the assumption of statistical independence of errors is actually valid, measurements should preferably be taken by different methods, with very different potential sources of error.

Traditionally, the validity of dietary questionnaire measurements is evaluated by comparison with repeated measurements of actual food consumption on a given day, obtained by food consumption records or 24-h diet recalls (both of these will also be referred to as 'reference' measurements (R), as, at a group level, they are often assumed to provide approximately unbiased measurements of intake, conditional on true intake level, i.e. $R = T + \varepsilon_R$). Using this traditional design, the correlation coefficient ρ_{QT} is estimated by:

1. calculating the crude coefficient of correlation between questionnaire measurements and the individuals' averages of their replicate reference measurements (food consumption records or 24-h diet recalls); and
2. correcting this crude coefficient for attenuation due to residual, within-subject variation in the individuals' average reference measurements, using a univariate analysis of variance to estimate within- and between-subject components of variance in the average reference measurements (Rosner & Willett, 1988).

As mentioned above, however, for the estimated correlation coefficient ρ_{QT} to be valid, crucial assumptions to be made are as follows:

- Random measurement errors are statistically independent between the questionnaire and the replicate reference measurements.
- Random measurement errors are statistically independent between the replicate reference measurements themselves.

Violation of the first assumption in the form of a positive correlation between errors of questionnaire and reference measurements will lead to an overestimation of the validity coefficient ρ_{QT}; on the other hand, a positive correlation between the errors of replicate reference measurements will lead

to an insufficient correction for attenuation effects, and thus to an underestimation of the correlation coefficient ρ_{QT}. If neither type of violation can be ruled out, it is difficult to predict whether the correlation coefficient ρ_{QT} will be overestimated or underestimated.

As mentioned above, biomarkers have the advantage that random errors can be assumed to be independent of those of more traditional types of dietary intake assessment. Thus, any correlation between marker and questionnaire or reference measurements is assumed to be due entirely to the fact that these measurements are associated with the same latent variable (true intake level of a given dietary constituent). Nevertheless, it is important to note that generally biomarkers will not have a perfect correlation with true intake level, even when the mean of many replicate measurements of the marker over time is taken (i.e. most markers have less than perfect validity as a measure for classifying individuals by habitual, true intake levels; see below). Therefore, the correlation between questionnaire measurements and markers, even though adjusted for attenuation effects due to random variation in the marker measurements themselves, can only be taken as an estimated *lower limit* for the correlation coefficient ρ_{QT}. A somewhat different use of biomarkers in dietary validity studies proposed more recently involves inferences based on simultaneous comparisons between questionnaire measurements (Q), 24-h food consumption records or 24-h recalls (R) and marker (M) (Plummer & Clayton, 1993a,b; Kaaks *et al.*, 1994a). For instance, from the a simple triangular comparison between single measures of Q, R and M, estimates of the correlations between each of the three types of measurements and individuals' true habitual intake levels (ρ_{QT}, ρ_{RT}, ρ_{MT}) can be obtained using a structural equations model (Kaaks *et al.*, 1994a) or, more directly, an elementary factor analysis model (Kaaks, 1997). The advantage of such a triangular comparison is that independence between random errors may need to be assumed only between the three *different* types of measurement (which indeed may have relatively independent sources of error in practice), and that no such assumption is needed between replicate measurements obtained by the same method (replicate weighed food records or replicate 24-h diet recalls). If it is felt that some level of positive correlation between random errors of questionnaire measurements (Q) and food consumption records or recalls (R) cannot be excluded (but that the random errors of markers would be independent of those of measurements Q and R), then the correlation coefficient ρ_{QT} estimated from the triangular comparison between measurements Q, R and M can at least be interpreted as an estimated *upper limit* of the true value of ρ_{QT} (Kaaks, 1997).

So far, substudies to evaluate the validity of dietary questionnaire measurements have usually been conducted either on an external study group of subjects not involved in the main epidemiological (cohort) study, or on a relatively small subsample of main study participants (nested within a cohort). Thus, since only a negligible number of cases with a specific disease would occur in the subsample, additional reference measurements (food consumption records, 24-h recalls) or additional measurements based on a biomarker could not be used as additional predictors of disease relative risk. However, when biological specimens are collected from all members of a cohort study, and are available for later biochemical analyses, markers can be measured in the specimens of subjects who subsequently develop disease and in individuals who have remained disease-free (e.g. using a nested case–control or a case–cohort design). In such situations, a triangular comparison between questionnaire measurements (Q), marker (M) and disease status (D) may in principle allow the estimation of relative risks unbiased by the random errors in either type of exposure measurement. However, the statistical methods for the analysis of studies on the association between diet and disease risk, using biomarkers as an additional measurement of dietary exposure, still require further investigation.

Sources of variation, reproducibility and validity

Intervening factors

The interpretation of a biomarker as a measure of an individual's (relative) intake level of a given food component obviously relies on the assumption that within- or between-subject variations in true intake levels of a given compound of interest are a primary determinant of within- and between-subject variations in the marker. Variations in the measured level of a marker of diet may, however, have many other exogenous or endogenous deter-

minants which may affect the absorption, distribution over body compartments or tissues, metabolism or excretion of a given compound. For example, active smoking, as well as exposure to environmental smoke, have been reported to be associated with decreases in plasma levels of vitamin C (Schectman *et al.*, 1989; Tribble *et al.*, 1993; Tappia *et al.*, 1995) or β-carotene (Stryker *et al.*, 1988; Saintot *et al.*, 1994) that could not be explained (only) by differences in diet, and which may be due to increased oxidation of these antioxidant vitamins. Another example, concerning the validity of biomarkers of fat-soluble compounds such as carotenoids, retinol or vitamin E, is the simultaneous intake of dietary fats, which enhance their intestinal absorption. Endogenous metabolic factors include mechanisms of homeostatic regulation of specific nutrient levels in blood plasma (e.g. homeostatic mechanisms controlling the intestinal absorption and circulating plasma levels of iron; homeostatic mechanisms controlling plasma levels of retinol, by secretion of retinol—bound to retinol-binding protein—from the liver), or factors that affect the distribution of a compound over body compartments and different tissues (e.g. vitamin C in plasma and in white blood cells). Other factors may be related to the metabolism and/or elimination of compounds. For instance, the levels of adducts to lymphocyte DNA of many potentially toxic substances, such as *N*-nitroso-compounds formed in the stomach from dietary precursors, or heterocyclic amines formed in meat during cooking, can be co-determined by genetic or phenotypic polymorphisms in the activity of enzymes that play a role in the detoxification and elimination of such compounds, or in the formation of reactive intermediates that can form covalent bonds with DNA. These intervening factors may cause considerable variation in the marker that is unrelated to a dietary factor of interest. This additional random variation ('random' in the sense that it is not predictable from the variations in diet that the marker is supposed to reflect) will attenuate the association between markers and true dietary intake levels.

Within a restricted range of intake levels, the dose–response relationship between intake levels of a given dietary constituent and a biochemical marker is often approximately linear, and intervening factors may cause mainly 'random' variations in the marker that are unrelated to true intake level. Over wider ranges of intake, however, homeostatic mechanisms may cause significant deviations from linearity. For example, for plasma retinol (vitamin A) concentrations, dose–response curves may be quasi-linear at habitual intake levels that are so low as to induce a retinol-deficient (or borderline deficient) state. At higher habitual intake levels, however, similar to those in most well-nourished populations, the dose–response curve flattens off to a practically constant level that is under homeostatic control by the liver. Clearly, in the flatter parts of this type of dose–response curve, the association between marker and intake level will be weaker, or the association may even be practically absent (see example of retinol in Table 1).

Reproducibility, validity and reliability

Two useful notions in describing the quality and usefulness of a biomarker of diet are its *reproducibility* (also called *precision*) and its *validity*. High reproducibility implies that a measure of a biomarker leads to very similar results for biological specimens collected on two independent occasions from the same individuals. A practical measure of the reproducibility is the correlation between the replicate measurements.

Validity means that a marker actually measures what it is intended to. As mentioned earlier, most markers cannot be translated into a measure of daily intake on an absolute scale, but at the very best are only correlated with intake level; that is, the marker can only be used to provide a relative classification of individuals by high or low dietary intake levels. Therefore, we shall use the term validity here in a somewhat restricted sense, to indicate that the average of an infinite (or very large) number of replicate measurements of the marker (on independent occasions) has a correlation of (close to) 1.0 with the individuals' true intake levels. Furthermore, we shall call a measurement of a marker of diet 'reliable' when it combines a high validity with a high reproducibility.

Perfect validity of a marker of diet implies that variations that are uncorrelated with true intake level ('random errors') are also statistically independent between replicate measurements taken on the same individual (after a sufficiently long time interval). When the validity is less than perfect,

part of the between-subject variation in the marker must reflect determinants (known or unknown) other than the dietary factor of interest. For example, plasma levels of a fat-soluble compound such as vitamin E may be influenced, not only by vitamin E intake, but also by plasma concentrations of low density lipoproteins, which, in turn, may depend on the type and amounts of fat consumed, on the degree of obesity, or on genetic factors causing between-subject variations in lipid metabolism.

Reproducibility is generally seen as an initial practical criterion for the selection of a marker of diet as a potential measure of exposure—by definition a poorly reproducible marker also has low reliability, irrespective of what the marker was supposed to measure. Nevertheless, if random variations in the marker are only weakly correlated between replicate measurements, the overall reliability may be increased substantially by taking the average of replicate measurements of the marker on the same individual. Inversely, when the random errors of replicate measurements are highly correlated (i.e. the validity is low), the reliability of a marker may not be improved much by taking the average of multiple measurements per individual.

In practice, the reproducibility of a biomarker will be a function of the stability of individuals' dietary intake level of a given compound (i.e. the ratio of within-subject to between-subject variations in intake level), as well as of its metabolic turnover rate. The turnover rate may depend largely on whether the compound is mainly fat-soluble or water-soluble, as well as on the type of medium (e.g. plasma, urine, saliva, blood cells or other tissue) chosen for the measurement. For example, after intake of a high dose of vitamin C, a water-soluble compound, a relatively large proportion will be excreted in the urine within a very short time. Fat-soluble compounds, on the other hand, can accumulate more in the body. The day-to-day variation of vitamin C concentrations in lymphocytes is lower than that of vitamin C levels in blood plasma. Thus, if the vitamin C concentrations in lymphocytes can be taken as a weighted average of varying plasma levels in the days or weeks before a blood sample is taken, the vitamin C concentrations in lymphocytes may be a more reliable marker of average intake levels over a longer time period. (Note, however, that results from several studies have shown that the reproducibility of measurements of plasma vitamin C levels, and the correlation of plasma levels with dietary intake measurements, can also be quite high (Bingham *et al.*, 1995); these high correlations suggest that individuals' dietary intake levels of vitamin C may be relatively stable and, hence, that a fast metabolic turnover may have a relatively minor impact on the reliability of this marker). Similarly, the day-to-day variability of fatty acid composition is higher for plasma triglycerides (especially in non-fasting plasma) than for plasma phopholipids, cholesteryl esters and erythrocyte membranes, and is almost null in subcutaneous adipose tissue biopsies, where the half-life of fatty acid turnover is estimated at 18–24 months.

A question of practical interest is whether a more reproducible marker—e.g. the fatty acid composition of an adipose tissue biopsy, rather than that of plasma triglycerides—may generally also be expected to be more valid. This would be true if the metabolic factors that cause the increased stability over time primarily smoothen the effects of day-to-day variations in true intake level, but do not induce new between-individual variations that are independent of the individuals' habitual intake levels. In reality, this may not always be the case. For example, from a physiological point of view, the fatty acid composition of (non-fasting) plasma triglycerides, which consist mainly of fats absorbed immediately after a meal, should reflect very well the composition of the fats just ingested. As a measure of the average composition of fats consumed in the long term, however, this marker may be relatively unreliable, as there can be substantial day-to-day variation in the fatty acid composition of foods consumed. Inversely, turnover rates of fats are low in adipose tissue, and the fatty acid composition of such adipose tissue biopsies is a highly reproducible measurement, at least over intervals of less than 1 year. Nevertheless, whereas some types of fatty acid from diet will be absorbed and finally stored in adipose tissue without modification and in proportion to their presence in plasma triglycerides, much of other fatty acids may be modified before or upon storage, and part of the fatty acids found in adipose tissue may actually have been synthesized endogenously from carbohydrates. Therefore, although more stable over time, the fatty acid composition of adipose tissue may only be a reliable marker of some (not all)

types of fatty acid in the diet (e.g. the linoleic acid content of adipose tissue shows moderately high correlations with linoleic acid intake from diet, whereas for palmitic acid or stearic acid this correlation tends to be lower: van Staveren *et al.*, 1986; London *et al.*, 1991).

Studies to evaluate the reliability (ranking capacity) of markers as a measurement of diet have generally been based on a simple comparison (correlation) with traditional dietary intake assessments (see Table 1 for examples). For some markers, e.g. carotenoids and vitamin E, correlations with questionnaire measurements have been reported to be rather low (Table 1; see also reviews by Willett, 1990b and van't Veer *et al.*, 1993), whereas the correlation can be as high as 0.6 or 0.8 for vitamin C or linoleic acid (van Staveren *et al.*, 1986; Bingham *et al.*, 1995; Riboli *et al.*, 1997). Adjustments for attenuation bias due to time-related, intra-individual variations in markers—especially those based on a concentration of a given compound in tissues or body fluids, rather than on a balance between intake and excretion—often do not lead to great improvements in the correlation with the dietary intake assessments, which suggests that the validity of these markers may be limited. Obviously, however, a less than perfect correlation between markers and other measurements of dietary intake (e.g. from a questionnaire) may also be due to errors in the latter type of measurement. The use of latent variable methods for the analysis of validity studies based on comparisons between replicate measurements of a biomarker, questionnaire measurements of habitual diet and/or replicate measurements of actual food consumption on given days may help to obtain more accurate estimates of the validity and precision of markers.

Validity and confounding

Besides being additional sources of variation in the levels of markers, exogenous and endogenous intervening factors may also be independent predictors of the risk of developing a given disease under study. For example, high plasma levels of fat-soluble compounds—such as β-carotene, vitamin E or DDT metabolites—may reflect high dietary intake levels of these compounds, but they may also reflect high concentrations of low-density and very low-density lipoproteins (LDLs, VLDLs) (Hunter, 1990, 1993). High plasma levels of LDLs and VLDLs are associated with a nutritional/metabolic profile (e.g. marked by insulin resistance) that is often found in obese individuals and which may be a risk factor of disease independently of, and possibly in an opposite direction to, vitamin E. It has been stated that, in this type of situation, the intervening factors may be seen as (positive or negative) confounders of the observed marker–disease relationship. In many of these situations, however, this view may be overly simplistic or even incorrect.

If the intervening factors mainly cause random variation in the marker, unrelated to the level of intake of a given dietary component of interest, while at the same time affecting disease risk, then the marker may not only reflect differences in intake levels of the dietary component, but partly will also be a reflection of the effects of the intervening factors; in this situation, the *interpretation* of the marker–disease relationship as reflecting uniquely a relationship between intake level and disease would be simply *invalid*. In a slightly more complicated case, intervening factors causing variation in both the markers and disease risk may also have some correlation with the dietary variable that a marker was actually supposed to measure. For instance, smokers may consume less fruit and vegetables and have a lower vitamin C intake (Schectman *et al.*, 1989; Subar *et al.*, 1990; Ross *et al.*, 1995), while, independently, inhalation of tobacco smoke may also lower plasma vitamin C levels by increasing its oxidation (Schectman, 1989; Tribble *et al.*, 1993; Tappia *et al.*, 1995). Furthermore, smoking augments the risk of developing many chronic diseases, including cancer. In this case, the relationship between disease risk and *intake* of vitamin C is indeed confounded by smoking, in the sense that intake levels of vitamin C could be found to be inversely associated with disease risk even when vitamin C intake was unrelated to the etiology of disease. To the extent that the marker is interpreted purely as a measurement of vitamin C intake level, the estimated relationship between *marker* and disease risk must then also be considered to be confounded by smoking. Again, however, as part of the variation in the marker may actually reflect the vitamin C-lowering effects of smoking rather than vitamin C

intake, one can argue that the interpretation of the marker purely as a measurement of vitamin C intake level is simply invalid. To avoid such ambiguities in interpretation, it seems conceptually more appropriate to consider the marker as an *intermediate* variable that is determined by dietary variables of interest together with various intervening factors, and at the same time as a determinant of disease risk. Under this alternative view, the relationship between vitamin C intake and the plasma concentration of vitamin C as an intermediate (outcome) variable is confounded by smoking, but not necessarily the relationship between disease risk and plasma vitamin C as an indicator of vitamin C status, which could actually be a valid reflection of the relationship between the biological availability of vitamin C and disease risk.

Summary and concluding remarks

Biomarkers as a measurement of dietary intake level can be useful when estimated intake levels of foods obtained from questionnaires or from records or short-term recalls of food consumption cannot be translated reliably into intake levels of specific compounds, for instance because of a lack of accurate food composition tables. In addition, even when dietary intake levels can be assessed with reasonable accuracy by means of a questionnaire, or by keeping records of actual food consumption, a marker can be a useful complementary measurement given that its random errors are likely to be truly independent of those of the first two categories of measurement. Thus, markers can be useful, especially in dietary validity studies, to estimate the average magnitude of systematic and random errors of dietary exposure measurements, as obtained in an epidemiological study. Knowledge of the average magnitude of such errors allows correction for bias in relative risk estimates.

A limitation of most markers as indicators of diet is that, at the very best, they can only provide an estimate of relative ranking of individuals by intake level, but cannot be translated into absolute intakes on a valid scale. This limitation is generally present when the marker is based on measurement of a concentration of a compound, rather than on the recovery over time of a compound in excretion products, and will be shared by many biomarkers that are currently being developed (e.g. biomarkers for intake of bioflavonoids, indoles or phytoestrogens from certain plant foods, or for intake of aromatic amines formed during cooking of meat or fish). In validity sub-studies, as well as in full-scale epidemiological studies of diet–disease relationships, this limitation can be overcome if another type of measurement is available (e.g. weighed food consumption records) which does provide estimates of intake level on an approximately valid scale. In theory, a few markers that can be translated into correctly scaled, absolute daily intake levels (e.g. 24-h urinary nitrogen excretion to estimate daily protein intake) could also be used as a reference measurement for *calibration* substudies. The latter can be used to adjust for bias in relative risks estimated for absolute, quantitative differences in dietary intake level (rather than between different quantile levels of the intake distribution) (Kaaks *et al.*, 1995), and can improve the comparability of dietary intake measurements between different study cohorts (Kaaks *et al.*, 1994b; Plummer *et al.*, 1994).

As described in this chapter, there can be many applications of biomarkers as indicators of diet in epidemiological studies. Nevertheless, some words of caution may be needed against overenthusiasm in favour of markers as the solution to the problems of mismeasurement of dietary intake levels by more traditional methods (questionnaires, food consumption records, short-term recalls). Results from many studies suggest that correlations between biomarkers and true intake levels of specific dietary constituents are generally not substantially higher (in fact, they are often lower) than those for the more traditional types of measurement, and even the reproducibility of markers can be relatively low. The rather low reliability of many markers may be due partly to a comparatively large intra-individual variation in true intake level over time and relatively fast metabolic turnover rates of given dietary compound, or may partly be explained by the effects of intervening variables that affect the validity of the marker as a measure of diet. Another aspect that may explain the low reproducibility of a marker are errors in measurements due to inaccuracy of laboratory techniques. Because of random measurement errors, and irrespective of the type of application of markers in

nutritional epidemiology—as an independent predictor of disease risk, or as an additional measurement of intake level in combination with more traditional measurements of diet—the statistical power of studies using markers will be low unless they include very large numbers of subjects.

Strong caution must also be exercised against overinterpretation (unjustified extrapolation) of the results of epidemiological studies based on markers, in that it may not be entirely clear which type of intake variable is specifically represented by a marker. For example, in several observational (epidemiological) studies, lung cancer risk was found to be associated inversely with plasma levels of β-carotene (ATBC, 1994; Omenn *et al.*, 1994), and this seemed to confirm results showing similar inverse associations with intake levels of β-carotene as measured by dietary questionnaires. Nevertheless, in the Finnish ATBC trial and in the CARET trial, supplementation with β-carotene (at supraphysiological doses) led to a significant increase, instead of the expected decrease, in lung cancer risk. These conflicting results suggest that the decrease in lung cancer risk observed in epidemiological studies may have been attributed wrongly to β-carotene *per se*, and that high estimated intake levels of β-carotene or high measured plasma concentrations were actually indicators of other protective factors, e.g. associated with a high intake of vegetables and fruits (although an alternative interpretation of these studies may be that β-carotene at physiological doses does actually confer some protection against lung cancer, but that supraphysiological doses are harmful).

This example shows that, in spite of a greater apparent specificity as a measurement of the relative intake level of a given dietary constituent, biomarkers in many (if not most) cases may only be interpreted as indicators of more complex dietary risk factors. Indeed, there is increasing awareness that it may be more appropriate to search for *patterns* of dietary risk factors associated with increased risk. In future studies, different markers combined may be used as indicators ('markers' in the true sense) of such more complex patterns.

Acknowledgements

We thank Dr David Hunter, Harvard School of Public Health, for reviewing this manuscript and for his helpful comments.

References

Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group (ATBC) (1994) The effect of vitamin E and beta-carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.*, 330, 1029–1035

Ascherio, A., Stampfer, M.J., Colditz, G.A., Rimm, E.B., Litin, L. & Willett, W.C. (1992) Correlations of vitamin A and E intakes with the plasma concentrations of carotenoids and tocopherols among American men and women. *J. Nutr.*, 122, 1792–1801

Berg, J.P., Glattre, E., Haldorsen, T., Hostmark, A.T, Bay, I.G., Johansen, A.F. & Jellum, E. (1994) Longchain serum fatty acids and risk of thyroid cancer: a population-based case–control study in Norway. *Cancer Causes Control*, 5, 433–439

Berrino, F., Muti, P., Micheli, A., Bolelli, G., Krogh, V., Sciajno, R., Pisani, P., Panico, S. & Secreto, G. (1996) Serum sex hormone levels after menopause and subsequent breast cancer. *J. Natl Cancer Inst.*, 88, 291–296

Bingham, S.A. & Cummings, J.H. (1985) Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am. J. Clin. Nutr.*, 42, 1276–1289

Bingham, S.A. & Cummings, J.H. (1986) Creatinine and PABA as markers for completeness of collection of 24-hour urine samples. *Hum. Nutr. Clin. Nutr.*, 40, 473–476

Bingham, S.A., Cassidy, A., Cole, T.J., Welch, A., Runswick, S.A., Black, A.E., Thurnham, D., Bates, C., Khaw, K.T., Key, T.J. & Day, N.E. (1995) Validation of weighed records and other methods of dietary assessment using the 24-hour urine nitrogen technique and other biological markers. *Br. J. Nutr.*, 73, 531–550

Bingham, S.A., Murphy, J., Waller, E., Runswick, S.A., Neale, G., Evans, D. & Cummings, J.H. (1992) Paraamino benzoic acid in the assessment of completeness of 24-hour urine. Collections from hospital outpatients and the effect of impaired renal function. *Eur. J. Clin. Nutr.* 1992, 46, 131–135

Cameron, M.E. & van Staveren, W.A. (1988) *Manual on Methodology for Food Consumption Studies*, Oxford, Oxford University Press

Campbell, D.R., Gross, M.D., Martini, M.F.C. Grandits, G.A., Slavin, J.L & Potter, J.D. (1994) Plasma carotenoids as biomarkers of vegetable and fruit intake. *Cancer Epidemiol. Biomarkers Prev.*, 3, 493–500

Coates, R.J., Eley, J.W., Block, G., Gunter, E.W., Sowell, A.L., Grossman, C. & Greenberg, R.S. (1991) An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income black women. *Am. J. Epidemiol.* 134, 658–671

de Klerk, N.H., English, D.R. & Armstrong, B.K. (1989) A review of the effects of random measurement error on relative risk estimates in epidemiological studies. *Int. J. Epidemiol.*, 18, 705–712

Donovan, U.M. & Gibson R.S. (1995) Iron and zinc status of young women aged 14 to 19 years consuming vegetarian and omnivorous diets. *J. Am. Coll. Nutr.*, 14, 463–472

Enger, S.M., Longnecker, M.P., Shikany, J.M., Swenseid, M.E., Chen, M., Harper, J.M. & Haile, R.W. (1995) Questionnaire assessement of intake of specific carotenoids. *Cancer Epidemiol. Biomarkers Prev.*, 4, 201–205

Feunekes, G.I.J., Van Staveren, W.A., De Vries J.H.M., Burema, J. & Hautvast J.G.A.J.(1993) Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol 1-3. *Am. J. Clin. Nutr.*, 58, 489–496

Gann, P.H., Hennekens, C.H., Sacks, F.M., Grodstein, F., Giovannucci, E.L. & Stampfer, M.J. (1994) Prospective study of plasma fatty acids and risk of prostate cancer. *J. Natl Cancer Inst.*, 86, 281–286

Garland, M., Willett, W.C., Manson, J.E. & Hunter D.J. (1993) Antioxidant micronutrients and breast cancer. *J. Am. Coll. Nutr.*, 12, 400–411

Guengerich, F.P. (1995) Influence of nutrients and other dietary materials on cytochrome P-450 enzymes. *Am. J. Clin. Nutr.* 61, 651S–658S

Hebert, J.R., Hurley, T.G., Hsieh, J., Rogers, E., Stoddard, A.M., Sorensen, G. & Nicolosi, R.J. (1994) Determinants of plasma vitamins and lipids: the working well study. *Am. J. Epidemiol.* 140, 132–147

Hunter, D. (1990) Biochemical indicators of dietary intake. In: Willett, W., ed., *Nutritional Epidemiology*, New York, Oxford University Press, pp. 375–418

Hunter, D.J. & Kelsey, K.T. (1993) Pesticide residues and breast cancer: the harvest of a silent spring? *J. Natl Cancer Inst.*, 85, 598–599

Hunter, D.J., Rimm, E.B., Sacks, F.M., Stampfer, M.J., Colditz, G.A., Litin, L.B. & Willett, W.C. (1992) Comparison of measures of fatty acid intake by subcutaneous fat aspirate, food frequency questionnaire, and diet records in a free-living population of US men. *Am. J. Epidemiol.*, 135, 418–427

Jacques, P.F., Sulsky, S.I., Sadowski, J.A., Phillips, C.C., Rush, D. & Willett, W.C. (1993) Comparison of micronutrient intake measured by a dietary questionnaire and biochemical indicators of micronutrient status. *Am. J. Clin. Nutr.*, 57, 182–189

Kaaks, R. (1997) Biomarkers as an additional measurement in studies on the accuracy of dietary questionnaire measurements. *Am. J. Clin. Nutr.* (in press)

Kaaks, R., Plummer, M., Riboli, E., Estève, J, & van Staveren, W. (1994b) Adjustment for bias due to errors in exposure assessments in multicenter cohort studies on diet and cancer: a calibration approach. *Am. J. Clin. Nutr.*, 59, 245S–250S

Kaaks, R., Riboli, E., Estève, J., van Kappel, A.L. & van Staveren W.A. (1994a) Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat. Med.*, 13, 127–142

Kaaks, R., Riboli, E. & van Staveren, W. (1995) Calibration of dietary intake measurements in prospective cohort studies. *Am. J. Epidemiol.*, 142, 548–556

Kabuto, M., Imai, H., Yonezawa, C., Neriishi, K., Akiba, S., Kato, H., Suzuki, T., Land, C.E. & Blot, W.J. (1994) Prediagnostic serum selenium and zinc levels and subsequent risk of lung and stomach cancer in Japan. *Cancer Epidemiol. Biomarkers Prev.*, 3, 465–469

Kardinaal, A.F.M., van't Veer, P., Brants H.A.M., van den Berg,H., van Schoonhoven, J. & Hermus R.J.J. (1995) Relations between antioxidant vitamins in adipose tissue, plasma, and diet. *Am. J. Epidemiol.*,141, 440–450

London, S.J., Sacks, F.M., Caesar, J., Stampfer, M.J., Siguel, E. & Willett, W.C. (1991) Fatty acid composition of subcutaneous adipose tissue and diet in postmenopausal US women. *Am. J. Clin. Nutr.*, 54, 340–345

London, S.J., Stein, E.A., Henderson, I.C., Stampfer, M.J., Wood, W.C., Remine, S., Dmochowski, J.R., Robert, N.J. & Willett, W.C. (1992) Carotenoids, retinol, and vitamin E and risk of proliferative benign breast disease and breast cancer. *Cancer Causes Control*, 3, 503–512

Ma, J., Folsom, A.R., Shahar, E. & Eckfeldt, J.H., for the Atherosclerosis Risk in Communities (ARIC) Study Investigators (1995) Plasma fatty acid composition as an indicator of habitual dietary fat intake in middle-aged adults. *Am. J. Clin. Nutr.*, 62, 564–571

Matkovic, V., Ilich, J.Z., Andon, M.B., Hsieh, L.C., Tzagournis, M.A., Lagger, B.J. & Goel, P.K. (1995) Urinary calcium, sodium, and bone mass of young females. *Am. J. Clin. Nutr.*, 62, 417–425

O'Donnell, M.G., Nelson, M., Wise, P.H. & Walker, D.M. (1991) A computerized diet questionnaire for use in diet health education 1. Development and validation. *Br. J. Nutr.*, 66, 3–15

Omenn, G.S., Goodman, G.E., Thornquist, M.D., Balmes J., Cullen, M.R., Glass, A., Keogh, J.P., Keyskens, F.L., Valanis, B., Williams, J.H., Barnhart, S. & Hammar, S. (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New Engl. J. Med.*, 334, 1150–1155

Plummer, M. & Clayton, D. (1993a) Measurement error in dietary assessment: an investigation using covariance structure models. Part I. *Stat. Med.*, 12, 925–935

Plummer, M. & Clayton, D. (1993b) Measurement error in dietary assessment: an investigation using covariance structure models. Part II. *Stat. Med.*, 12, 937–948

Plummer, M., Clayton, D. & Kaaks, R. (1994) Calibration in multi-centre cohort studies. *Int. J. Epidemiol.*, 23, 419–426

Porrini, M., Gentile, M.G. & Fidanza, F. (1995) Biochemical validation of a self-administered semi-quantitative food-frequency questionnaire. *Br. J. Nutr.* 74, 323–333

Preobrazhenskaya, M.N., Bukhman, V.M., Korolev, A.M. & Efimov, S.A. (1993) Ascorbigen and other indole-derived compounds from Brassica vegetables and their analogs as anticarcinogenic and immunomodulating agents. *Pharmacol. Ther.*, 60, 301–313

Riboli, E. & Kaaks, R. (1996) The EPIC project: rationale and study design. *Int. J. Epidemiol.* (in press)

Riboli, E., Rönnholm, H., Saracci, R. (1987) Biological markers of diet. *Cancer Surv.*, 6, 685–718

Riboli, E., Elmståhl, S., Saracci, R., Gullberg, B. & Lindgärde, F. (1997) The Malmö food study: validity of two dietary assessment methods for measuring nutrient intake. *Int. J. Epidemiol.* (in press)

Roidt, L., White, E., Goodman, G.E, Wahl, P.W., Omenn, G.S., Rollins, B. & Karkeck, J.M. (1988) Association of food frequency questionnaire estimates of vitamin A intake with serum vitamin A levels. *Am. J. Epidemiol.*, 128, 645–654

Romieu, I., Stamfer, M.J., Stryker, W.S., Hernandez, M., Kaplan, L., Sober, A., Rosner, B. & Willett, W.C. (1990) Food predictors of beta-carotene and alpha-tocopherol: validation of a food frequency questionnaire. *Am. J. Epidemiol.*, 131, 864–876

Rosner, B. & Willett, W.C. (1988) Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am. J. Epidemiol.*, 127, 377–386

Ross, M.A., Crosley, L.K., Brown, K.M., Duthie, S.J., Collins, A.C., Arethur, J.R. & Duthie, G.G. (1995) Plasma concentrations of carotenoids and antioxidant vitamins in Scottish males: influences of smoking. *Eur. J. Clin. Nutr.*, 49, 861–865

Saintot, M., Astre, C., Scali, J. & Gerber, M. (1995) Within-subjects seasonal variation and determinants of inter-individual variations of plasma beta-carotene. *Int. J. Vitam. Nutr. Res.*, 65, 169–174

Schectman, G., Byrd, J.D. & Gruchow, Hl.W. (1989) The influence of smoking on vitamin C status in adults. *Am. J. Public Health*, 79, 158–162

Schoeller, D.A. (1988) Measurement of energy expenditure in free-living humans by using doubly labeled water. *J. Nutr.*, 118, 1278–1289

Shaw, N.S., Chin, C.J. & Pan, W.H. (1995) A vegetarian diet rich in soybean products compromises iron status in young students. *J. Nutr.*, 125, 212–219

Sinha, R., Block, G. & Taylor, P.R. (1992) Determinants of plasma ascorbic acid in a healthy male population. *Cancer Epidemiol. Biomarkers Prev.*, 1, 297–302

Sinha, R., Frey, C.M., Kammerer, W.G., McAdams, M.J., Norkus, E.P. & Zeigler, R.G. (1994) Importance of supplemental vitamin C in determining serum ascorbic acid in controls from a cervical cancer case–control study: Implications for epidemiological studies. *Nutr. Cancer*, 22, 207–217

Sinha, R., Patterson, B., Mangels, A.R., Levander, O.A., Gibson, T., Taylor, P.R. & Block, G. (1993) Determinants of plasma vitamin E in healthy males. *Cancer Epidemiol. Biomarkers Prev.*, 2, 473–479

Sinha, R., Rothman, N., Brown, E.D., Salmon, C.P., Knize, M.G., Swanson, C.A., Rossi, S.C., Mark, S.D., Levander, O.A. & Felton, J.S. (1995) High concentrations of the carcinogen 2-amino-1-methyl-6-phenylimidazo-[4,5-b]pyridine (PhlP) occur in chicken but are dependent on the cooking method. *Cancer Res.*, 55, 4516–4519

Skog, K., Steineck, G., Augustsson, K. & Jagerstad, M. (1995) Effect of cooking temperature on the formation of heterocyclic amines in fried meat products and pan residues. *Carcinogenesis*, 16, 861–867

Snook, J.T., Palmquist, D.L., Moxon, A.L., Cantor, A.H. & Vivian, V.M. (1983) Selenium status of a rural (predominantly Amish) community living in a low-selenium area. *Am. J. Clin. Nutr.*, 38, 620–630

Stähelin, H.B., Gey, K.F., Eichholzer, M., Ludin, E., Bernasconi, F., Thurneysen, J. & Brubacher, G. (1991) Plasma antioxidant vitamins and subsequent cancer mortality in the 12-year follow-up of the prospective Basel Study. *Am. J. Epidemiol.*, 133, 766–775

Steinmetz, K.A. & Potter, J.D. (1991) Vegetables, fruit, and cancer. II. Mechanisms. *Cancer Causes Control*, 2, 427–442

Stryker, W.S., Kaplan, L.A., Stein, E.A., Stampfer, M.J., Sober, A. & Willett, W.C. (1988) The relation of diet, cigarette smoking, and alcohol consumption to plasma beta-carotene and alpha-tocopherol levels. *Am. J. Epidemiol.*, 127, 283–296

Subar, A.F., Harlan, L.C. & Mattson, M.E. (1990) Food and nutrient intake differences between smokers and non-smokers in the US. *Am. J. Public Health*, 80, 1323–1329

Swanson, C.A. Longnecker, M.P., Veillon, C., Howe, S.M., Levander, O.A., Taylor, P.R., McAdam, P.A., Brown, C.C., Stampfer & M.J., Willett, W.C. (1990) Selenium intake, age, gender, and smoking in relation to indices of selenium status of adults residing in a seleniferous area. *Am. J. Clin. Nutr.*, 52, 858–862

Tappia, P.S., Troughton, K.L., Langley-Evans, S.C. & Grimble, R.F. (1995) Cigarette smoking influences cytokine production and antioxidant defences. *Clin. Sci. Colch.*, 88, 485–489

Tjønneland, A., Overad, K., Thorling, E. & Ewertz, M. (1993) Adipose tissue fatty acids as biomarkers of dietary exposure in Danish men and women. J. *Clin. Nutr.*, 57, 629–633

Toniolo, P.G., Levitz, M., Seleniuch-Jacquotte, A., Banerjee, S., Koenig, K.L., Shore, R.E., Strax, P. & Pasternack, B.S. (1995) A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J. Natl Cancer Inst.*, 87, 190–197

Ullrey, D.E. (1981) Selenium in the soil-plant-food chain. In: Spallholtz, J.E. *et al.*, eds., *Selenium in Biology and Medicine*, Westport, CT, AVI Publishing, pp. 176–191

Tribble, D.L., Guiliana, L.J. & Fortmann, S.P. (1993) Reduced plasma ascorbic acid concentrations in nonsmokers regularly exposed to environmental tobacco smoke. *Am. J. Clin. Nutr.*, 58, 886–890

van den Brandt, P.A., Goldbohm, R.A., van't Veer, P., Bode, P., Dorant, E., Hermus, R.J. & Sturmans, F. (1993) A prospective cohort study on selenium status and the risk of lung cancer. *Cancer Res.*, 53: 4860–4865

van Staveren, W.A., Deurenberg, P., Katan, M.B., Burema, J., de Groot, L.C. & Hoffmans, M.D. (1986) Validity of the fatty acid composition of subcutaneous fat tissue microbiopsies as an estimate of the long-term average fatty acid composition of the diet of separate individuals. *Am. J. Epidemiol.*,123, 455–463

van't Veer, P., Wielen, V.D., Kok, F.J., Hermus, R.J.J. & Sturmans, F. (1990) Selenium in diet, blood, and toenails in relation to breast cancer: A case–control study. *Am. J. Epidemiol.*, 131, 987–994

van't Veer, P., Kardinaal, A.F., Bausch-Goldbohm, R.A. & Kok, F.J. (1993) Biomarkers for validation. *Eur. J. Clin. Nutr.*, 47 (Suppl. 2), S58–S63

Wattenberg, L.W. (1992) Inhibition of carcinogenesis by minor dietary constituents. *Cancer Res.*, 52, 2085S–2091S

Wild, J., Seller, M.J., Schorah, C.J. & Smithells, R.W. (1994) Investigation of folate intake and metabolism in women who have had two pregnancies complicated by neural tube defects. *Br. J. Obstet. Gynaecol.*, 101, 197–202

Willett, W.C. (1990a) Food frequency methods. In: Willett, W., ed., *Nutritional Epidemiology*, New York, Oxford University Press, pp. 69–91

Willett, W.C. (1990b) Reproducibility and validity of food frequency questionnaires. In: Willett, W., ed., *Nutritional Epidemiology*, New York, Oxford University Press, pp. 92–126

Williams, D.R. & Bingham, S.A. (1986) Sodium and potassium intakes in a representative population sample: estimation from 24 h urine collections known to be complete in a Cambridgeshire village. *Br. J. Nutr.*, 55, 13–22

Witschi, J.C. (1990) Short-term dietary recall and recording methods. In: Willett, W., ed., *Nutritional Epidemiology*, New York, Oxford University Press, pp. 52–68

Yang, G., Zhou, R., Yin, S., Gu, L., Yan, B., Liu, Y., Liu, Y. & Li, X. (1989) Studies of safe maximal daily dietary selenium intake in a seleniferous area in China. I. Selenium intake and tissue selenium levels of the inhabitants. *J. Trace Elem. Electrolytes Health Dis.*, 3, 77–87

Yong, L.C., Forman, M.R., Beecher, G.R., Graubard, B.I., Campbell, W.S., Reichman, M.E., Taylor, P.R., Lanza, E., Holden, J.M. & Judd, J.T. (1994) Relationship between dietary intake and plasma concentrations of carotenoids in premenopausal women: application of the USDA-NCI carotenoid food-composition database. *Am. J. Clin. Nutr.*, 60, 223–230

Zheng,W., Blot, W.J., Diamond, E.L., Norkus, E.P., Spate, V., Morris, J.S. & Comstock, G.W. (1993) Serum micronutrients and the subsequent risk of oral and pharyngeal cancer. *Cancer Res.* 53, 795–798

Corresponding author

R. Kaaks

Unit of Nutrition and Cancer, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 8, France

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Biomarkers for biological agents

N. Muñoz and F.X. Bosch

Biomarkers of exposure to biological agents have proved to be extremely useful in establishing causal associations between infections and human cancer, as well as in tracing the natural history of the relevant agents. This chapter will focus on biomarkers for biological agents currently recognized as causally associated with various human cancers after evaluation at the IARC. These are human papillomavirus (HPV), hepatitis B (HBV) and C virus (HCV) and *Helicobacter pylori*.

A prototype of nucleic-acid-based biomarkers is detection of HPV DNA. It measures the presence of type-specific DNA at a given point in time. PCR-based assays are considered the method of choice for epidemiological investigations. The test requires collection of exfoliated cells or biopsies. Specimens can be kept stored at –20°C for long periods of time. DNA degradation is low. Some of the limitations of the marker for cancer epidemiology lie in the fact that HPV DNA infections are often transient, especially in young women. In repeated measurements, HPV DNA may fluctuate, but the reasons for this are unknown.

Antigens and antibodies from the HBV and HCV can be viewed as prototypes for serological biomarkers. For HBV, there are markers able to distinguish between past and persistent infections. HBV DNA detection in sera further refines the assessment of exposure. Standardized serological assays are available and widely used in developing countries. For HCV antibodies, serological assays are standardized and widely available. RNA detection in sera by PCR is under development. A limitation of the currently available assays is the type and subtype variation of HCV by geography, which requires further research and standardization. In low-risk populations, numerous false-positive results arise and confirmatory tests are required.

Biomarkers of exposure to biological agents have proven to be extremely useful in establishing causal associations between infectious agents and human cancer, as well as in tracing the natural history of the relevant agents. This chapter will focus on biomarkers for biological agents currently recognized as causally associated with various human cancers after evaluation at the International Agency for Research on Cancer (IARC). These are human papillomavirus (HPV) (IARC, 1995), hepatitis B (HBV) and C virus (HCV) (IARC, 1994a) and *Helicobacter pylori* (IARC, 1994b).

HPV and cervical cancer

Over 75 types of HPV have so far been identified. One of the most common methods of virus detection, namely the isolation of virus in tissue culture, is not available for HPVs because HPVs cannot be propagated in culture. The presence of the virus can be inferred from clinical, cytological, histological or serological findings, but the most accurate methods are based on the demonstration of HPV DNA in cells and, to a lesser extent, on the identification of viral proteins in the cells. Serological assays, currently under development, could be of great help in epidemiological studies as they might be able to assess cumulative lifetime exposure or at least recent, as opposed to just current, infection. The colposcopic, cytological and histological diagnoses of HPV are not very reliable as they lack sensitivity and specificity (Barrasso, 1992; Sherman *et al.*, 1994) and as such have very limited use in epidemiological studies. Likewise, the identification of viral proteins such as HPV capsid antigens in infected cells is also of limited epidemiological value as they can only be detected in the case of productive HPV infection. These methods will not be discussed here. We will concentrate on methods for HPV DNA detection and on serological markers of HPV.

Methods for HPV DNA detection

HPV DNA detection methods have been developed most extensively for genital tract HPVs, because these infections have been linked to cervical cancer. Of the 30–40 HPVs that infect the genital tract, some types (e.g. HPV types 16, 18, 45, 31 and 33)

are closely associated with cervical cancers, whereas others (e.g. HPV types 6, 11, 42, 43, 44) are rarely found in cancers. Therefore, in clinical and epidemiological investigations, HPV diagnosis aims at identification of specific types of HPVs and at separating cancer-associated HPVs from low-risk HPVs.

Two types of HPV DNA hybridization methods are currently used—with and without amplification. In the former, targeted viral sequences are first amplified by polymerase chain reaction (PCR) and then identified by hybridization assays. Most authors would consider PCR methods as the current standard for epidemiological studies. These methods have been reviewed recently (Gravitt & Manos, 1992; Walboomers *et al.*, 1994) and summarized in IARC (1995).

Serological assays are still under development. Currently, ELISA (enzyme-linked immunoabsorbent assay) antibody assays using virus-like particles as the antigen appear to be the most promising and have raised high expectations. However, the antibody response is low-titred and not detectable in all patients with documented infections. Likewise, at present, antibodies to E6 and E7 proteins are the most promising markers of HPV-associated invasive cancer.

Logistics of sample collection, processing and storage

Exfoliated cells can be collected from easily accessible sites such as uterine cervix, penis, bladder or oral cavity, but collection from other organs such as oesophagus, stomach or lung requires invasive procedures, i.e. endoscopy. Various methods are used to collect exfoliated cells: swabs, scrapes, brushings, lavages and urinary sediments. Tissue specimens are obtained from biopsies or surgical specimens and therefore only from subjects with disease. Ethical restriction on taking biopsies from healthy individuals precludes the use of tissue specimens in control subjects.

Whatever the specimen, the major drawback of PCR-based methods is the risk of contamination as a result of the high sensitivity of these methods. The entire process (from sample collection to the PCR assay) is susceptible to contamination. Thus, it is recommended that strict negative controls be strategically interspersed during every step of the assay (Gravitt & Manos, 1992).

Sources of variability

High interlaboratory variability has been reported for FISH, Southern blot and the Hybrid Capture™ assay. There is a significant interlaboratory variation in detection and typing of HPVs by Southern hybridization. In a study of 40 clinical samples, pairwise agreement between four laboratories ranged from 66 to 97% for HPV detection, and from 77 to 96% for typing of positive specimens (Brandsma *et al.*, 1989).

In a study of interlaboratory variation in Hybrid Capture results in three laboratories, specimens were tested with probe pools A and B. Kappa values in interlaboratory pairwise comparisons for positivity to either group A or group B ranged from 0.61 to 0.83. Of specimens that were positive for group B by the reference standard, 74% were positive for group B in all three laboratories. The interlaboratory correlations of HPV quantitative data for probe B types ranged from 0.60 to 0.90. Probable false-positive results were occasionally encountered, in less than 3% of the tests (Schiffman *et al.*, 1995).

Interlaboratory variation for two PCR-based methods using L1 consensus primers has been shown to be small (Hsing *et al.*, 1996).

HPV DNA detection in a single specimen measures the presence of the virus at a given point in time, which is clearly different from assessing lifetime exposure to HPV. At present, there are no reliable methods (hybridization or serological) to measure HPV lifetime exposure or to distinguish transient from persistent HPV infections other than repeated HPV DNA detection.

HPV types are defined by genomic analysis and therefore represent genotypes. At present, an HPV genome is described as a new HPV type if the nucleotide sequences of its *E6*, *E7* and *L1* genes (about one-third of the genome) differ by more than 10% from those of any previously described HPV type. Although this definition was established arbitrarily, it appears to be sufficient to define the five natural taxonomic units. HPV types differing by 2–10% are called subtypes and those differing by less than 2% are called variants. Short-term longitudinal studies, such as those aimed at determining the natural history of HPV infections, ideally should rely on the identification of variants, but genome sequencing is too time-consuming and costly to be applied to large-scale epidemiological studies.

Applications of the biomarker in epidemiological studies

Biomarker as the end-point

Natural history of HPV infections. Prevalence surveys of HPV DNA in various populations have shown that most HPV infections are latent or subclinical, as most individuals in whom HPV DNA is detected do not develop signs or symptoms. Short-term longitudinal studies have shown that most HPV infections are transient, especially at younger ages, and that persistent infections, although less common, pose a higher risk of pre-neoplastic lesions.

Transmission studies. Results from various longitudinal studies indicate that genital HPVs are transmitted primarily through contact with infected tissues during sexual intercourse. Infections of the oropharynx and larynx may be related to orogenital contact, and mother to infant transmission has been proposed to explain recurrent respiratory papillomatosis in infants and children.

Use of biomarker in case–control/prospective cohort studies of cervical cancer

The epidemiological evidence indicating that some HPV types are human carcinogens at least for the uterine cervix is compelling and it is based on an impressive and largely consistent set of case series and case–control studies and some cohort studies (IARC, 1995).

Case series from many populations around the world have shown the presence of HPV DNA in the majority of high-grade CIN lesions (CIN II–III) and cervical carcinomas. In the largest study, in which about 1000 invasive cancer specimens were analysed in a single laboratory using a reliable PCR-based assay, a prevalence of 93% was found, with no significant geographical variation in overall positivity. HPV 16 was present in 50% of the specimens, HPV 18 in 14%, HPV 45 in 8% and HPV 31 in 5% (Bosch *et al.*, 1995).

Case–control studies. Although a large number of studies have been published, only 10 fulfil the epidemiological requirements of a case–control study and have used accurate HPV DNA assays. Six were on high-grade CIN lesions and four were on invasive cervical cancer. In all of them, a strong association with certain HPV types (16, 18, 31, 33, 35) was detected, with odds ratios (ORs) ranging from 15 to 73 for high-grade CIN, and 15 to 46 for invasive cervical cancer (Muñoz & Bosch, 1996).

Serological assays using various early or late HPV antigens have shown that, in general, antibody responses are higher in women with cervical neoplasia than among control women, but the magnitudes of the ORs are lower than those based on HPV DNA detection (Dillner, 1994).

Cohort studies. Several cohort studies have been reported, but only three have used high-grade CIN as the end-point and accurate methods of HPV DNA detection. All of them suggest that persistent infection with high-risk HPV types precedes the development of high-grade CIN (Muñoz & Bosch, 1996). Recent observations suggest that persistent HPV infection, particularly with a high viral load, is a predictor of progression to advanced CIN, although the truly prospective estimates were low (Ho *et al.*, 1995).

An association between HPV and other anogenital cancers (cancers of the anus, penis and vulva) has been reported in a few case–control studies and has been suggested for cancers of the upper aerodigestive tract (IARC, 1995).

Limitations of the evidence. Case–control studies suffer from inherent temporal ambiguity concerning exposure and disease outcome. Thus, the higher prevalence of HPV DNA among cases than among controls could be interpreted in two ways. If we assume that a single measurement of HPV DNA is a good marker of chronic persistent infection, HPV DNA detected at recruitment of cases and controls could be regarded as a marker of an HPV infection that preceded the cancer development. Alternatively, HPV DNA could be more readily detected in tumoral cells than in normal cells or could be a marker of an opportunistic infection with HPV.

Direct evidence in support of the first possibility can only be derived from long-term follow-up studies, and few such studies are now available. In addition, indirect evidence may be obtained from the trend of HPV DNA prevalence by time since last sexual intercourse, because the major route of transmission is sexual. Data from studies in Spain and Colombia show a stable high rate of HPV DNA positivity both in women with cervical cancer who reported being sexually active at the time of the interview and in women who had their last sexual

intercourse many years before entry into the study (Muñoz *et al.*, 1992; Bosch *et al.*, 1993).

The possibility of enhanced detectability in tumoral cells is unlikely, because the HPV DNA prevalence in precursor lesions (CIN II–III) is as high as in invasive cervical cancer. Running contrary to the argument that HPV is an opportunistic infection is a great deal of laboratory data indicating that DNA and transcripts of specific HPV types are usually detected in tissue specimens from cervical cancer and its precursor lesions, and that high-risk HPVs are able to immortalize human cells and their oncoproteins are able to interfere with the functions of negative cellular regulators.

Results from case–control studies and our international prevalence survey of HPV DNA in invasive cervical cancer indicate that over 95% of these tumours can be attributed to certain HPV types. The fact that only a small minority of the persistent HPV infections progress eventually to cancer indicates that there must be other factors or cofactors that increase the progression rate to malignancy.

Two types of cofactors may be of importance:

- Host factors that modulate the effect of HPV, such as genetic factors [HLA or major histocompatibility complex (MHC) haplotypes], genetic or induced immunosuppression, endogenous hormonal factors, reflected in the associations with high parity detected in our studies, as well as early age at first sexual intercourse, which could be regarded as a surrogate of early age at first HPV infection.
- Exogenous factors. In the studies in Spain, Colombia and Brazil, only long-term use of oral contraceptives emerged as a cofactor among HPV-positive women. However, our observations need to be confirmed in other populations and in larger studies.

Future directions

Epidemiological studies aimed at investigating the natural history of HPV infections (prevalence, incidence, persistence) and their role in various diseases rely on PCR-based assays. The two most widely used methods, which amplify a region of the *L1* gene of genital HPVs (MY09–MY11 and GP5+/GP6+), need to be formally compared with each other and also with the recently described PCR-based assay that amplified *E6* sequences of genital HPVs (Lungu *et al.*, 1995). Standardized reagents for these three methods are needed. Although the two methods that amplify a region of the *L1* gene are able to detect HPV DNA in over 90% of invasive cervical cancers, recent studies indicate the possibility that neither of these systems will readily detect some of the HPV intratypical variants of 12 genital HPVs (Stewart *et al.*, 1996).

PCR-based assays in an ELISA format to distinguish high-risk HPVs (16, 18, 31, 33, 35, 39, 45, 51, 52, 54, 56 and 58) from low-risk HPVs (6, 11, 34, 40, 42, 43 and 44) have been developed (Jacobs *et al.*, 1995) and are being commercialized. These methods will be of great help in screening programmes and clinical management of HPV-associated lesions and as a first step in epidemiological investigations.

The ELISA assay to detect antibodies to virus-like particles was, until recently, considered the most promising serological assay for the detection of HPV-specific antibodies. However, this assay appears to have low sensitivity and specificity, as antibodies are not detected in 25–50% of women with HPV 16-positive cervical cancers, and cross-reactivity with other HPV types has not yet been excluded. In addition, the antibody response is low-titred and it declines with time. Accurate serological assays which are able to distinguish past transient infections from persistent infections need to be developed.

Biomarkers for *Helicobacter pylori*

Helicobacter pylori is a spiral or slightly curved Gram-negative bacterium that is uniquely adapted to survive the acidic environment of the stomach of humans and non-human primates. *H. pylori* strains are genetically heterogeneous, can be cultured in a microaerophilic environment, are sensitive to most antibiotics *in vitro* and are characterized by very strong urease activity. *H. pylori* is recognized as the main cause of gastritis and peptic ulcer, and has also been associated with stomach cancer and gastric lymphoma. Colonization of the gastric mucosa and subsequent development of gastritis are dependent on bacterial factors such as motility, strong urease activity and specific adherence to gastric epithelium.

Methods for detection of *H. pylori* and some sources of variability

Methods for detection of *H. pylori* infection have been summarized in IARC (1994b). In brief, these

use as biological specimens gastric biopsies, smears on gastric material, gastric juice, breath tests and serology.

Gastric biopsies are collected before treatment from both the antrum and the corpus with standard forceps and can be cultured after (1) placing them in saline (analysis within 4 h); (2) placing them in transport medium (analysis after up to 24 h); or (3) freezing them at –70°C or in liquid nitrogen (delayed analysis).

Methods based on biopsies include: (1) a rapid urease test; (2) histological examination stained with the standard haematoxylin-eosin or the modified Giemsa stain; (3) bacteriological tests including Gram staining or culture; (4) PCR with primers that encode urease, 16S ribosomal RNA or specific genes of pathogenic relevance, such as the *cagA* gene.

The techniques used for gastric biopsy specimens can also be used with gastric juice samples. Techniques based on faecal specimens and cultures from dental plaque and saliva are still in an early stage of development.

Urea breath tests were developed taking advantage of the ability to hydrolyse urea by the strong urease of *H. pylori*. In the urea breath test, urea labelled with $^{13}CO_2$ is fed and subsequently eliminated in the breath. Breath samples are collected before and 30 min after absorption of labelled urea and analysed for $^{13}CO_2$ by mass spectrometry. High levels indicate infection. Similar tests involve the use of ^{14}C-urea, as $^{14}CO_2$ can be measured easily with a scintillation counter, but some concern has been expressed over the use of a radioactive isotope. Low-dose tests are being developed to overcome this problem.

A variety of serological assays have been used, including complement fixation, haemagglutination, bacterial agglutination, immunofluorescence and ELISA, which is currently the technique of choice.

Various ELISA commercial kits have been compared and large variations have been observed, especially when the assay was repeated on different days (Feldman & Evans 1995).

Currently, the most accurate 'gold standard' for a positive is probably culture supplemented with a histology or biopsy urease test. In epidemiological studies involving children, the breath test is usually used as gold standard.

Applications of the biomarker in epidemiological studies

Biomarker as the end-point

Seroprevalence surveys have shown that *H. pylori* prevalence is highest in developing countries, where 80–90% of the population may be infected by early adulthood, whereas in developed countries, in the age group 25–34 years, the proportion of antibody-positives ranges from 10–60%. The prevalence increases gradually with age; it is higher in lower socio-economic groups and when crowding in childhood is reported. Transmission occurs from person to person, but the exact mode has not been elucidated.

Use of biomakers in case–control/prospective cohort studies of gastric cancer

The epidemiological evidence linking *H. pylori* to stomach cancer and gastric lymphoma was summarized in IARC (1994b) as follows:

- Ecological studies. Six out of eight studies in which the prevalence of *H. pylori* was correlated with concurrent or earlier incidence or mortality from cancer of the stomach show no correlation between the two variables, and in the other two studies a weak correlation was observed.
- Case series. In 10 case series of stomach cancer in which *H. pylori* was detected by histology, the proportion of positives ranged from 43 to 83%, and it was over 90% in two case series of gastric lymphomas.
- Case–control studies. At least 10 case–control studies have addressed the association between prevalence of *H. pylori* antibodies and gastric cancer with inconsistent results. In low-risk countries for gastric cancer three out of four studies reported significantly increased risk with ORs around 2–3. In high-risk countries, four out of six studies have been negative (Muñoz & Pisani, 1994).
- Cohort studies. Five case–control studies nested within cohorts have been reported, and in three of them an increased risk of gastric cancer was observed, with prospective RRs ranging from 2.8 to 6.0, while the other two failed to show an association (Muñoz & Pisani, 1994).

Although the working group at the IARC judged the available evidence as sufficient to classify

H. pylori as carcinogenic to humans, it seems to us that the quality of most of the epidemiological studies evaluated is such that biases and confounding cannot be ruled out (Muñoz & Pisani, 1994).

A recognized limitation of current serological methods of *H. pylori* detection for epidemiological studies is that the amount of *H. pylori* antibodies may be affected by the presence of stomach cancer. It has been observed that *H. pylori* does not colonize gastric mucosa that has undergone intestinal metaplasia. Thus, patients with gastric cancer showing extensive areas of intestinal metaplasia in the non-cancerous mucosa might have a low bacterial load and, consequently, low titres of antibodies which may be difficult to detect.

Future directions

It is possible that *H. pylori* strains prevalent in high-risk populations for stomach cancer differ antigenically from those prevalent in low-risk areas. Thus, development of serological assays using antigens derived from local *H. pylori* strains and validated in the populations under study should be encouraged.

In addition, two types of *H. pylori* strains are now recognized. Type I strains possess *cag A* (cytotoxin-associated gene A), a gene encoding a high-molecular-weight immunodominant antigen and *vac A* (vacuolating cytotoxin gene). Type II strains do not have *cag A* or *vac A* genes. Type I strains of *H. pylori* have been associated with an increased risk of gastric cancer, but also of duodenal ulcer and atrophic gastritis (Figura *et al.*, 1989; Fox *et al.*, 1992; Blaser *et al.*, 1995; Kuipers *et al.*, 1995). ELISA serological assays to detect antibodies to cag A and vac A protein have been developed (Fox *et al.*, 1992; Blazer *et al.*, 1995), but have not yet been validated. These assays, once validated, may be of value in epidemiological studies addressing the association between *H. pylori* and stomach cancer.

Hepatitis B viral markers

Methods for HBV detection and interpretation of results

HBV infections—acute, persistent or resolved—are detected on the basis of serological assays for viral antigens and antibodies, as well as detection of nucleic acids. In serum specimens, the most commonly used markers in epidemiological studies related to liver cancer are the hepatitis B surface antigen (HBsAg), the hepatitis B surface antibody (HBsAb), the hepatitis B core antibody (HBcAb), the hepatitis B e antigen (HBeAg) and the hepatitis B e antibody (HBeAb).

PCR technology is rapidly evolving and several methods and variants are being proposed to increase the accuracy of detecting HBV markers and to introduce quantitative measurements of viral DNA (see recent examples in Gunther *et al.*, 1995; Maia *et al.*, 1995; Ranki *et al.*, 1995). Other methods are being developed to amplify sequences of HBV and HCV simultaneously(Nedjar *et al.*, 1994; Hu *et al.*, 1995). If validated, these methods will make a remarkable contribution to field epidemiology. PCR also facilitates analysis of the sequence of the amplified genomes. PCR followed by sequencing has also been used for subtyping HBV and for characterizing and identifying HBV mutants.

HBV is almost unique for cancer epidemiology in that typical patterns of serological markers in HBV infection have been described and are useful in interpreting HBV exposure (see Table 1).

Logistics of sample collection

Blood samples are among the easiest specimens to collect in field studies. Preprocessing and long-term storage for HBV markers detection requires little more than standard laboratory equipment and –20°C freezers. Several of the assays described above are standardized, commercially available and routinely used in blood banks and hospital laboratories. Technology transfer to developing countries in Africa and Asia has been successfully achieved.

Sources of variation across studies

Sources of variation between studies include the following:

1. The laboratory assays used.
2. The lack of detail on the histological classification of cases.
3. The use of hospital controls with different stringency criteria for inclusion and exclusion. This is particularly relevant with respect to the inclusion of patients with other liver diseases as controls.
4. In several instances, the HBsAg prevalence among hospital controls is used to estimate the population attributable fraction (AF). It has been reported that hospital patients have higher HBsAg prevalence rates than the general

Table 1. Typical serological patterns in HBV infection

Infection status	HBsAg	Anti-HBc IgM	Anti-HBc Total	HBeAg	Anti-HBe	Anti-HBs
Acute infection[a]	+	+	+	+	−	−
Chronic infection with high levels of viral replication	+	−	+	+	−	−
Chronic infection with low levels of viral replication[b]	+	−	+	−	+	−
Recovery from acute infection before development of anti-HBs	−	+	+	−	+	−
Low titre; possible false-positive	−	−	+	−	−	−
High titre; possible 'low-level carrier'	−	−	+	−	+	−
Recovery from acute infection, indicating immunity	−	−	+	−	+	+
Vaccine response[c]	−	−	−	−	−	+
Susceptible to HBV infection	−	−	−	−	−	−

[a]Reactivated chronic disease may have this pattern with sensitive anti-HBc IgM assays. [b]Some patients may be seronegative for HBeAg and anti-HBe. [c]In unvaccinated individuals, a high titre may represent immunity or be non-specific; low titres are often non-specific. (From IARC, 1994a.)

population (Maynard et al., 1989; Vall Mayans *et al.*, 1990), thus underestimating the OR and perhaps overestimating or underestimating the AF.

Evidence is now accumulating that HBV carriers might not have HBsAg in their sera. These are usually HBcAb-positives without HBsAg who are shown to harbour HBV DNA in the sera and in the liver by the recently applied DNA amplification techniques (PCR), monoclonal antibodies and Southern blot assays (Nakajima *et al.*, 1989). The use of these methods in epidemiological studies might increase the accuracy of our estimates of the relative risk for hepatocellular carcinoma (HCC) among HBV carriers and the fraction of the disease that is attributable to HBV.

Contamination remains a major difficulty of studies based on PCR methods. Negative and positive control samples, including reaction mixtures without DNA, should be analysed in each test (Dusheiko *et al.*, 1992; Seelig *et al.*, 1992). One large validation study using a panel of serum specimens was organized in 39 laboratories to evaluate variation in the detection of HBV DNA using non-standardized PCRs. Of 43 sets of results, 10 (23.3%) correctly identified the entire panel of specimens (seven HBV-DNA-positive specimens, five HPV-DNA-negatives and two dilution series). Nineteen (44.2%) had some false-positive or false-negative results with the undiluted specimens, and 40% had errors with the diluted specimens (Quint *et al.*, 1995).

Although the biological matrix of interpretation of the HBV profile is fairly consistent, in some situations the serological results are affected by the coexistence of additional factors.

Among subjects with inflammatory hepatopathies, patients on haemodialysis, intravenous drug addicts, homosexuals and also in a fraction of spontaneous blood donors, anti-HBc may appear as the sole marker of HBV exposure (reviewed by Levine *et al.*, 1994). Studies have shown that, in these patients, HBsAg is often immunocomplexed (80% in inflammatory hepatopathies, 40% in patients on haemodialysis, 63% in intravenous drug addicts), escaping detection by standard tests. Moreover, among patients with complexed HBsAg, 48% were HBV-DNA-positive by nested PCR, compared with 19% among patients without complexed HBsAg. These results suggest that some of the patients with anti-HBc may only develop an active HBV infection (Joller-Jemelka *et al.*, 1994).

Epidemiological studies using HBV biomarkers

Biomarkers as the end-point

The majority of the epidemiological studies on HBV and HCC have been based on the one-time detection of HBsAg, HBsAb and HBcAb. Some studies have also evaluated HBeAg and HBeAb. The

persistence of HBsAg in the serum following infection can be viewed as a surrogate marker of the risk of developing HCC and an intermediate end-point for natural history studies.

Likewise, the persistence of detectable levels of HBsAb can be interpreted as a measure of protection conferred by natural infection or following vaccination. Some examples and principal results of studies conducted using these HB markers include HBV transmission studies and studies to evaluate HBV vaccine efficacy.

The five basic HBV markers have been extensively used to investigate HB transmission from HBsAg-carrier mothers to newborns, between sexual couples or between siblings. The detection of HBeAg/Ab has been useful in ascertaining the increased risk of transmission from mothers to newborns if the mother is a carrier of the HBeAg. Furthermore, recent work from the Gambia has shown that it is unlikely that bed-bug bites are responsible for HB transmission (Vall Mayans *et al.*, 1990).

Vaccine efficacy has been assessed using short-term follow-up studies in which the development of HBsAb in vaccinated children is used as one of the end-points. Prevention of the development of the HBsAg-carrier state (i.e. by comparing with a control group) provides surrogate estimates of the prevention of HCC to be expected in the future. These follow-up studies are also useful in estimating the duration of the protection afforded by HB vaccination

Newly developed biomarkers may modify our current estimates of the risk related to HBV. One relates to the existence of HBV mutants that may escape detection using conventional methods; the other source of HBV exposure misclassification relates to the HBV infections with low replication level in which PCR techniques are required to detect minute quantities of HBsAg or HBV DNA.

Use of biomarkers in case–control/prospective cohort studies of HCC

Hepatitis B is remarkable in that several antibody profiles can be described and attributed with some exceptions to specific stages of the host/viral interactions (Table 1). These markers were sufficient to establish a strong link between the persistence of HBsAg in serum and the risk of HCC. In brief, epidemiological studies have been able to demonstrate:

1. The geographical variation in the prevalence of HBsAg carriers and its correlation with the incidence of HCC (Szmuness, 1978).
2. The intrafamilial and intravillage clustering of HBV infections (Barrett et al., 1977; Whittle *et al.*, 1991).
3. The strong association of HBV and HCC in more than 40 case–control studies. HBsAg carriage is the strongest determinant of HCC. In some studies, the presence of HBcAb is also associated with the risk. In low-risk countries in Europe and the USA, the OR estimates range from 5 to 40 and the population AF ranges from 1 to 50%, with a median value around 20%. In high-risk areas in China and South-East Asia, ORs (10–30) and AFs (range 40–90%) tend to be higher.
4. The independence of HBV from other risk factors and the specificity of the association with liver cancer (largely HCC) as opposed to other cancers metastatic to the liver.
5. The risk of HCC among HBsAg carriers in several prospective cohort studies (Beasley *et al.*, 1981; McMahon *et al.*, 1990; Hall *et al.*, 1985). In some cohort studies, the risk of HCC was highest for HBsAg carriers who were also HBeAg carriers (Sakuma *et al.*, 1988). As in case–control studies, risk estimates for HBsAg and attributable fractions are high in high-risk countries such as Taiwan (Beasley *et al.*, 1981) and in Asian migrant populations in low-risk countries such as the USA (Sherman *et al.*, 1995). The risk of progression is low or very low in the native populations of low/intermediate-risk countries such as Canada (Villenueve *et al.*, 1994) and Italy (De Francis et al., 1993)
6. The efficacy of HBV vaccines in preventing the HBsAg-carrier state in all settings (McMahon *et al.*, 1987; Hsu *et al.*, 1988; Tsen *et al.*, 1991; Whittle *et al.*, 1991).
7. Interactions of HBV and aflatoxin (AF) in the causation of liver cancer. Epidemiological studies are increasingly using biomarkers of exposure to AF. Two of the most promising are urinary metabolites and adducts of AF and the mutation patterns of the p53 tumour suppressor gene.

In addition, HBV markers have been used to identify high-risk groups and in the evaluation of treatment modalities:

1. Regular screening of HBsAg carriers for early liver cancers has been attempted in high-risk populations. The results in terms of improving survival of the cases identified are not consistent (McMahon *et al.*, 1991; De Francis *et al.*, 1993; Villenueve *et al.*, 1994; Colombo 1995).
2. HBV vaccines are also under trial as a treatment for HBsAg carriers. In these studies, measurements of HBV load are used as biomarkers in the evaluation of therapy (Poll *et al.*, 1993).

The AFB_1-*N*-Gua adducts and other AF metabolites have been used in an ongoing cohort study in Shanghai, China (Ross *et al.*, 1992; Qian *et al.*, 1994). This study showed that seropositivity to HBsAg and the presence of one to four AF exposure markers in urine (AFB_1, AFP_1, AFM_1 and AFB_1-*N*-Gua) were strongly related to HCC (OR = 59.4; 16.6–212.0). There was also a moderate increase in risk for subjects exposed to AF who were HBsAg-negative (OR = 3.4; 1.1–10.0) (Qian *et al.*, 1994). Subsequently, it was indicated that the anti-HCV prevalence in the HCC cases found in the cohort was about 1%, and in the control group was about 0.2% (OR = 5.0; 0.3–79.9) (Yuan *et al.*, 1995). This study is so far the best evidence of an interaction between HBV and AF in the causation of human HCC. However, the number of HCC cases in which the interaction was explored is small (13 cases AF-positive and HBsAg-negative) and there is room for misclassification of cases in relation to their viral exposure (thus risk estimates become unstable). Better measurements of HBV exposure (i.e. detection of HBV DNA in serum by PCR) have shown that in areas where HBV is prevalent in South-East Asia, HBV DNA can be identified in up to 20% of the HBsAg-negative HCC cases (reviewed by Paterlini & Brechot, 1994) and in 30–32% of patients with chronic liver disease (Zhang *et al.*, 1993). The follow-up of this cohort is awaited with great interest.

The presence of AF-specific mutations in the genome of HCC cells was suggested by an ecological study in which HCC biopsies from patients in 14 countries were investigated. Among 72 HCC specimens from South Africa and the south-east coast of Asia (assumed to have been exposed to AF), 12 (17%) showed a specific G to T mutation at codon 249 of the *p53* gene. The mutation was not found in any of 95 HCC specimens from a miscellany of countries where AF exposure is rare (Ozturk, 1991). Other reports with limited numbers of cases indicated high rates (i.e. >50% of the *p53* codon 249 mutations) in HCC specimens from Quidong (Hsu *et al.*, 1991), Senegal (Coursaget *et al.*, 1993) and South Africa (Bressac *et al.*, 1991). Lower frequencies were reported from HCC in Thailand, where AF contamination of foods has been documented in the past (Hollstein *et al.*, 1993), and among Eskimos (De Benedetti *et al.*, 1995). The genetic changes associated with HCC are just beginning to be understood (Tabor, 1994).

These new biological markers may represent a real breakthrough in the field of HCC epidemiology. In particular, the new genetic markers can be determinant in quantifying the responsibility of AF as an independent cause of HCC and in evaluating the likely interactions with the hepatitis viruses in humans.

However, a word of caution should be raised concerning these pioneer studies in relation to: (1) their small sample size and limited methodology with respect to criteria for specimen inclusion; (2) inadequate adjustment of the correlations for exposures to other (viral and non-viral) risk factors at the individual level; (3) limited information on the sensitivity and specificity of the assays to assess proposed genetic markers; and (4) insufficient knowledge on the additional genetic changes associated with HCC development.

Future directions

In spite of the large number of studies addressing natural history issues, we still lack an explanation for the occurrence of a substantial fraction of HBV infections in most settings. Natural history studies should be pursued based upon the newer methods of HBV detection currently available.

PCR methods and new assays based on different regions of the viral genome (i.e. HBxAg and anti-HBx; Horiike *et al.*, 1991), pre-S antigens and antibodies (Itoh *et al.*, 1986; Coursaget *et al.*, 1990) may be useful in the future to increase precision in the quantification of the HBV involvement in the incidence of HCC. Studies on viral integration may contribute to the description of the mechanisms of carcinogenesis.

The coexistence of HBV and HCV, with or without HDV, may be greater than 10%. The mecanisms by which HCV seems to suppress HBV and

HDV replication remain unknown. The impact of multiple hepatitis infections on HCC incidence requires further research. Follow-up studies of haemophiliacs and drug addicts may provide unique opportunities for research.

Molecular techniques are identifying genetic changes related to exposure to liver carcinogens, such as aflatoxins (Ozturk, 1991), or to cell foci believed to be HCC precursors (for a review see Tabor, 1994; Hsia *et al.*, 1994).

Polymorphism of the MHC has been suggested as playing an important part in the establishment of the HBV persistent infection (Thursz *et al.*, 1995). Other work suggests that non-response to HBV vaccines may be related to HLA groups (A. Zuckerman, personal communication, 1996). More studies using these susceptibility markers are needed.

The investigation of HBV mutants is in constant progress. The clinical significance of some of these mutant viruses in the induction of HCC is unknown.

Increasing the precision of the estimates concerning HBV (and HCV) will allow better estimates of the role of other putative factors such as tobacco, alcohol and oral contraceptives. It should also help in understanding the role of liver cirrhosis and other chronic liver conditions as intermediate pre-neoplastic stages.

Hepatitis C viral infections

Methods for the detection of HCV infection and sources of variability

The detection of infection by HCV is based upon assays for viral antibodies and viral nucleic acids. Tests to detect exposure to HCV have been in constant development since first reported in 1989 (Kuo *et al.*, 1989). Second- and third-generation ELISA assays, confirmatory assays with recombinant immunoblot assays (RIBA 1-4) and quantitative tests based on PCR are now available.

HCV infection can also be assessed by detecting HCV RNA by reverse transcription (RT) and PCR, which is highly sensitive and has been used for early diagnosis. Quantitative PCR can be used to detect 5–30 molecules of synthetic HCV RNA (Hagiwara *et al.*, 1993).

Testing by PCR has become the standard for some workers. The results of these tests correlate well with the risk for transmitting post-transfusion hepatitis, with the results of second-generation anti-HCV assays and with liver histology, and are useful in monitoring the response of patients to interferon therapy. Nevertheless, reproducibility between laboratories, each using their own protocols, for detection of anti-HCV has been poor (Zaaijier *et al.*, 1993).

Well-controlled procedures for handling samples, extraction and purification of nucleic acids, avoidance of laboratory contamination and use of appropriate negative and positive controls are essential prerequisites for the PCR assay. Selection of primers from the highly conserved 5' noncoding region is also important for sensitivity and has allowed identification of a broad range of genotypes (Okamoto *et al.*, 1990).

Sequence variation analysis allowed the description of distinct types of HCV with some geographical variation in the type-specific prevalence (Dusheiko *et al.*, 1994). Six major types were recognized in 1994 (Simmonds *et al.*, 1994). However, these classifications are contingent on completing sequencing and it is expected that a recognized system of HCV classification will be available in 1996. Some commercial kits are available for HCV detection and typing (Bréchot & Thiers, 1996).

HCV RNA testing in mononuclear peripheral blood cells offers the possibility to detect HCV replication in patients who are HCV-RNA-negative in serum (Muratori *et al.*, 1994). The clinical value of these tests is uncertain, but it has been suggested that HCV replication in mononuclear cells can be responsible for the reactivation and acute episodes of HCV hepatitis during chronic infection and after transplantation (Qian *et al.*, 1992). These markers have not been used in epidemiological studies.

Logistics of sample collection

Epidemiological studies have been based on serological assays. The logistics required for blood collection in the field are similar to the description pertaining to HBV.

Applications of biomarker in epidemiological studies

Biomarkers as the end-point

The presence of HCV markers in serum can be interpreted as a surrogate measure of the risk of developing liver cirhosis and HCC. Most of the

available evidence refers to natural history studies and therapeutic intervention studies.

More than half of the cases of HCV exposure remain of obscure origin, in spite of intensive questioning. Person to person exposure is the most likely source, but the mechanism is poorly defined. Among haemodialysis patients, HCV contamination has been documented in wards with strict control of blood products with second-generation assays. DNA sequencing of the hypervariable region of the *E2* gene in PCR products identified clusters of closely related viruses, suggesting common sources other than blood (Allender *et al.*, 1995). Other studies based on PCR have shown the presence of HCV RNA in tears and other body fluids, suggesting a possible alternative source of exposure (Feutch *et al.*, 1994; Su *et al.*, 1994). HCV RNA titre in the serum correlates with the presence of the viral RNA in seminal fluid or saliva of infected males (Liou *et al.*, 1992) and with the rate of transmission from mothers to infants (Ohto *et al.*, 1994). These results suggest that serum HCV RNA titre is a marker of HCV infectivity.

The mode of acquisition of HCV is of relevance to progression to chronic hepatitis and liver cirrhosis. Several studies have shown that transfusion-acquired HCV conveys a poorer prognosis than both sporadic hepatitis C (Jové *et al.*, 1988; Mattson *et al.*, 1989) and hepatitis C following i.v. drug use (IDU) (Gordon *et al.*, 1993). Studies of viral load using serum titre of HCV RNA have shown that transfused patients had a higher viral load than both patients infected through IDU and health workers (Lau *et al.*, 1993). Other studies suggest that long-term outcomes of HCV infection may be predicted by the severity of the initial hepatic lesion postinfection (Schmid *et al.*, 1982; Tito *et al.*, 1990).

Although there are few long-term follow-up studies of HCV-infected individuals, current evidence suggests that few people resolve HCV infections spontaneously, and HCV RNA can be found by PCR in most instances (Alter *et al.*, 1992).

There is also evidence of familial clustering of HCV. Mother to child transmission has been documented. Sexual contacts of anti-HCV carriers are at a moderately high risk of HCV. Patients attending STD (sexually transmitted diseases) clinics show higher HCV prevalence than the general population, although sexual transmission seems to occur at a lower rate than for HIV or HBV.

Use of biomarker in case–control/prospective cohort studies of HCC

Epidemiological studies are largely consistent in showing a strong association between anti-HCV and HCC. The specific potential of each of the HCV types and their variants to induce HCC, as well as the potential for other factors from the host and the environment to interact with HCV in the origin of HCC, still requires further research.

More than 30 case series on anti-HCV prevalence in HCC, liver cirrhosis and chronic hepatitis have been reported. The range of HCV seropositivity in HCC cases from developed countries is 20–76%, with most studies in the 50–70% range. Apart from exceptions, the prevalence rates from studies that used early assays are not significantly distinct from those that used second-generation assays.

The coexistence of HBV and HCV markers is relatively rare among HCC patients. Studies based on second-generation anti-HCV assays are consistent in showing prevalence rates of dual infection among cases below 10%. Relatively high rates (7.7%) have been found in Italy (Stroffolini *et al.*, 1992) and Senegal (4.1%) (Coursaget *et al.*, 1992). These results are consistent with the hypothesis that a substantial fraction of the anti-HCV-positive subjects observed in the developed countries result from transfusions with HBsAg-screened blood products.

Cohort studies have shown an increased risk for HCC among anti-HCV carriers. It is estimated that about 50% of the HCV infections will lead to chronic liver disease, of which 20% will develop liver cirrhosis and perhaps 10% of these with cirrhosis will progress to HCC (IARC, 1994a). Some studies estimate the average time interval from post-transfusion HCV hepatitis to chronic liver disease, liver cirrhosis and HCC at around 10, 20 and 30 years, respectively (Kiyosawa *et al.*, 1984).

Case–control studies on HCC using first- and second-generation assays are numerous and largely consistent in showing elevated risk estimates among anti-HCV carriers. In Japan, a study of 91 HCC cases and 410 controls from the general population reported an OR of 52 (24–111) with an HCV AF of 60.5% (Tanaka *et al.*, 1991). In Italy, a study of 65 cases and 99 controls reported an OR of 27 with an AF of 77% (Stroffolini *et al.*, 1992). Increased risk estimates have occasionally been obtained even in areas where HBV is the

predominant etiological factor of HCC, e.g. in Guanxi, southern China (Okuno *et al.*, 1994) and in Korea (Lee *et al.*, 1993). In these two areas, the prevalence rates of anti-HCV in all HCC cases (positive and negative for HBsAg) were 5 and 17%, respectively, and in HBsAg-negative HCC cases they were 1.8 and 43%. In a series of 23 studies from different countries published in 1991–1993, the range of the reported ORs is 1.1–134. These estimates were statistically significant in 18 of the 23 studies (for a review see also IARC, 1994a). Other risk factors such as smoking and alcohol do not modify the strength of the association between HCV and HCC.

From case–control studies, risk estimates are consistently higher for individuals who are anti-HCV-positive and HBsAg-positive (e.g. see Chuang *et al.*, 1993). It has recently been estimated that the attributable fraction for both HBV and HCV is about 50% in developed countries and 90% in the developing world (Pisani *et al.*, 1996).

Future directions

Genome sequencing, completion of the phylogenetic tree and agreement on a standard typification of HCV would help in evaluating international variation in HCV types. Follow-up studies of HCV-infected individuals should provide information on the risk factors for progression to advanced liver disease. Natural history studies and population prevalence surveys will be useful in identifying background prevalence, routes of transmission and preventive strategies. More sensitive methods of HCV detection and, in particular, quantitative PCR systems should be used in natural history studies.

References

Allender, T., Gruber, A., Mojgan, N., Beyene, A., Söderström, T., Björkholm, M., Grillner, L. & Persson, M.A.A. (1995) Frequent patient-to-patient transmission of hepatitis C virus in a haematology ward. *Lancet*, 345, 603–607

Alter, M.J., Margolis, H.S., Krawczynski, K., Judson, F.N., Mares, A., Alexander, W.J., Hu, P.Y., Miller, J.K., Gerber, M.A., Sampliner, R.E., Meeks, E.L. & Beach, M.J. (1992) The natural history of community-acquired hepatitis C in the United States. *New Engl. J. Med.*, 327, 1899–1905

Barrasso, R. (1992) Colposcopic diagnosis of HPV cervical lesions. In: Muñoz, N., Bosch, F.X., Shah, K.V. & Meheus, A., eds, *The Epidemiology of Human Papillomavirus and Cervical Cancer* (IARC Scientific Publications No. 119). Lyon, International Agency for Research on Cancer, pp. 67–74

Barrett, D.H., Burks, J.M., McMahon, B., Elliott, S., Berquist, K.R., Bender, T.R. & Maynard, J.E. (1977) Epidemiology of hepatitis B in two Alaskan communities. *Am. J. Epidemiol.*, 105, 118–122

Beasley, R.P., Hwang, L.Y., Lin, C.C. & Chien, S.C. (1981) Hepatocellular carcinoma and hepatitis B virus: a prospective study of 22707 men in Taiwan. *Lancet*; ii, 1129–1133

Blaser, M.J., Pérez-Pérez, G.I., Kleanthous, H., Cover, T.L., Peek, R.M., Chyou, P.H., Stemmermann, G.N. & Nomura, A. (1995) Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.*, 55, 2111–2115

Bosch, F.X., Muñoz, N., de Sanjosé, S., Navarro, C., Moreo, P., Ascunce, N., Gonzalez, L.C., Tafur, L., Gili, M., Larrañaga, I., Viladiu, P., Daniel, R.W., Alonso de Ruiz, P., Aristizabal, N., Santamaria, M., Guerrero, E. & Shah, K.V. (1993) Human papillomavirus and cervical intraepithelial neoplasia grade III/carcinoma *in situ*: a case–control study in Spain and Colombia. *Cancer Epidemiol. Biomarkers Prev.*, 2, 415–422

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V. & The IBSCC Study Group (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J. Natl Cancer Inst.*, 87, 796–802

Brandsma, J.L., Burk, R.D., Lancaster, W.D., Pfister, H. & Schiffman, M.H. (1989) Inter-laboratory variation as an explanation for varying prevalence estimates of human papillomavirus infection. *Int. J. Cancer*, 43, 260–262

Bréchot, C. & Thiers, V. (1996) *Evaluation du kit de sérotypage du VHC* (Murex technical report), Murex

Bressac, B., Kew, M., Wands, J. & Ozturk, M. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature*, 350, 429–430

Chuang, W.L., Chang, W.Y., Lu, S.N., Lin, Z.Y., Chen, S.C., Hsieh, M.Y., Wang, L.Y., You, S.L. & Chen, C.J. (1993) The role of hepatitis C virus in chronic hepatitis B virus infection. *Gastroenterol. Jpn*, 28 (suppl. 5), 23–27

Colombo, M. (1995) Should patients with chronic viral hepatitis be screened for hepatocellular carcinoma? *Viral Hepatitis Rev.*, 1, 67–75

Coursaget, P., Buisson, Y., Bourdil, C., Yvonnet, B., Molinie, C., Diop, M.T., Chiron, J.P., Bao, O. & Diop-Mar, I. (1990) Antibody response to preS1 in hepatitis-B-virus-induced liver disease and after immunization. *Res. Virol.*, 141, 563–570

Coursaget, P., Leboulleux, D., Le Cann, P., Bao, O. & Coll-Seck, A.M. (1992) Hepatitis C virus infection in cirrhosis

and primary hepatocellular carcinoma in Senegal. *Trans. R. Soc. Trop. Med. Hyg.*, 86, 552–553

Coursaget, P., Depril, N., Chabaud, M., Nandi, R., Mayelo, V., Le Cann, P. & Yvonnet, B. (1993) High prevalence of mutations at codon 249 of the p53 gene in hepatocellular carcinomas from Senegal. *Br. J. Cancer*, 67, 1395–1397

De Benedetti, V., Welsh, J., Trivers, G., Harpster, A., Parkinson, A., Lanier, A., McMahon, B. & Bennett, W. (1995) P53 is not mutated in hepatocellular carcinomas from Alaska Natives. *Cancer Epidemiol. Biomarkers Prev.*, 4, 79–82

De Francis, R., Meucci, G., Vecchi, M., Tatarella, M., Colombo, M., del Ninno, E., Rumi, M.G., Donato, M.F. & Ronchi, G. (1993) The natural history of asymptomatic hepatitis B surface antigen carriers. *Ann. Intern. Med.*, 118, 191–194

Dillner, J. (1994) Antibody responses to defined HPV epitopes in cervical neoplasia. *Papillomavirus Report*, 5, 35–41

Dusheiko, G., Xu, J. & Zuckermann, A.J. (1992) Clinical diagnosis of hepatitis B infection: applications of the polymerase chain reaction. In: Becker, Y. & Darai, G., eds, *Frontiers of virology*. Berlin, Springer Verlag. pp. 67–85

Dusheiko, G., Schmilovitz-Weiss, H., Brown, D., McOmish, F., Yap, P.L., Sherlock, S., McIntyre, N. & Simmonds, P. (1994) Hepatitis C virus genotypes: an investigation of type-specific differences in geographic origin and disease. *Hepatology*, 19, 13–18

Feldman, R.A. & Evans, S.J.W. (1995) Accuracy of diagnostic methods used for epidemiological studies of Helicobacter pylori. *Alimen. Pharmacol. Therap.*, 9 (suppl. 2), 21–31

Feutch, H.H., Polywka, S., Zollner, B. & Laufs, R. (1994) Greater amount of HCV-RNA in tears compared to blood. *Microbiol. Immunol.*, 38, 157–158

Figura, N., Guglielmetti, P., Rossolino, A., Barberi, A., Cusi, G., Musmanno, R.A., Russi, M. & Quaranta, S. (1989) Cytotoxin production by Campylobacter pylori strains isolated from patients with peptic ulcers and from patients with chronic gastitis only. *J. Clin. Microbiol.*, 27, 225–226

Fox, J.G., Correa, P., Taylor N.S. Thompson, N., Fontham, E., Janney, F., Sobhan, M., Ruiz, B. & Hunter, F. (1992) High prevalence and persistence of cytotoxin-positive Helicobacter pylori strains in a population with high prevalence of atrophic gastritis. *Am. J. Gastroenterol.*, 87, 1554–1560

Gordon, S.C., Elloway, R.S., Long, J.C. & Dmuchowski, C.F. (1993) The pathology of hepatitis C as a function of mode of transmission: blood transfusion vs. intravenous drug use. *Hepatology*, 18, 1338–1343

Gravitt, P.E. & Manos, M.M. (1992) Polymerase chain reaction-based methods for the detection of human papillomavirus DNA. In: Muñoz, N., Bosch, F.S., Shah, K.V. & Meheus, A., eds, *The Epidemiology of Human Papillomavirus and Cervical Cancer* (IARC Scientific Publications No. 119). Lyon, International Agency for Researcho on Cancer, pp. 121–133

Gunther, S., Li, B.C., Miska, S., Kruger, D.H., Meisel, H. & Will, H. (1995) A novel method for efficient amplification of whole hepatitis B virus genome permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J. Virol.*, 69, 5437–5444

Hagiwara, H., Hayashi, N., Mita, E., Naito, M., Kasahara, A., Fusamoto, H. & Kamada, T. (1993) Quantitation of hepatitis C virus RNA in serum of asymptomatic blood donors and patients with type C chronic liver disease. *Hepatology*, 17, 545–550

Hall, A.J., Winter, P.D. & Wright, R. (1985) Mortality of hepatitis B positive blood donors in England and Wales. *Lancet*, 1, 91–93

Ho, G.Y., Burk, R.D., Klein, S., Kadish, A.S., Chang, C.J., Palan, P., Basu, J., Tachezy, R., Lewis, R. & Romney, S. (1995) Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *J. Natl Cancer Inst.*, 87, 1365–1371

Hollstein, M.C., Wild, C.P., Bleicher, F., Chutimataewin, S., Harris, C.C., Srivatanakul, P. & Montesano, R. (1993) P53 mutations and aflatoxin B_1 exposure in hepatocellular carcinoma patients from Thailand. *Int. J. Cancer*, 53, 51–55

Horiike, N., Blumberg, B.S. & Feitelson, M.A. (1991) Characteristics of hepatitis BX antigen, antibodies to Xantigen, and antibodies to the viral polymerase during hepatitis B virus infection. *J. Infect. Dis.*, 164, 1104–1112

Hsia, C.C., Di Bisceglie, A.M., Kleiner, D.E. & Tabor, E. (1994) Expression of hepatitis B surface and core antigens and transforming growth factor-a in "Oval Cells" of the liver in patients with hepatocellular carcinoma. *J. Virol.*, 43, 216–221

Hsing, A.W., Burk, R.D., Liaw, K.-L., Chen, C.-J., Zhang, T., Schiffman, M., Greer, C.E., You, S.-L., Hsieh, C.-Y., Huang, T.W., Wu, T.-C., O'Leary, T.J., Seidman, J.D. & Manos, M.M. (1996) Interlaboratory agreement in a polymerase chain reaction-based human papillomavirus DNA assay. *Cancer Epidemiol. Biomarkers Prev.*, 5, 483–484

Hsu, H.M., Chen, D.S., Chuang, C.H., Lu, J.C.F., Two, D.M., Lee, C.C., Lu, H.C., Cheng, S.H., Wang, Y.F., Wang Chen, C., *et al.* (1988) Efficacy of a mass hepatitis B vaccination program in Taiwan. *JAMA*, 260, 2231–2235

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J. &

Harris, C.C. (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature*, 350, 429–430

Hu, K.-Q., Yu, C.-H., Lee, S., Villamil, F.G. & Vierling, J. (1995) Simultaneous detection of both hepatitis B virus DNA and hepatitis C virus RNA using a combined one-step polymerase chain reaction technique. *Hepatology*, 21, 901–907

IARC (1994a) *IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans*. Vol. 59, *Hepatitis viruses*, Lyon, International Agency for Research on Cancer

IARC (1994b) *IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans*. Vol. 61, *Schistosomes, liver flukes and* Helicobacter pylori, Lyon, International Agency for Research on Cancer

IARC (1995) *IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans*. Vol. 64, *Human papillomavirus*, Lyon, International Agency for Research on Cancer

Itoh, Y., Takai, E., Ohnuma, H., Kitajima, K., Tsuda, F., Machida, A., Mishiro, S., Nakamura, T., Miyakawa, Y. & Mayumi, M. (1986) A synthetic peptide vaccine involving the product of the pre-S (2) region of hepatitis B virus DNA: protective efficacy in chimpanzees. *Proc. Natl Acad. Sci. USA*, 83, 9174–9178

Jacobs, M.V., de Roda Husman, A.M., van den Brule, A.J.C., Snijders, P.J.F., Meijer C.J.L.M & Walboomers, J.M.M. (1995) Group specific differentiation between high- and low-risk human papillomavirus genotypes by general primers-mediated polymerase chain reaction and two cocktails of oligonucleotide-probes. *J. Clin. Microbiol.*, 33, 901–905

Joller-Jemelka, H.I., Wicki, A.N., Grob, P.J. (1994) Detection of HBs antigen in "anti-HBc alone" positive sera. J *Hepatology*, 21, 269–272

Jové, J., Sánchez-Tapias, M., Bruguera, M., Mas, A., Costa, J., Barrera, J.M. & Rodes, J. (1988) Posttransfusional vs. sporadic non-A, non-B chronic hepatitis: a clinicopathological and evolutive study. *Liver*, 8, 42–47

Kiyosawa, K., Akahane, Y., Nagata, A. & Furuta, S. (1984) Hepatocellular carcinoma after non A non B posttransfusion hepatitis. *Am. J. Gastroenterol.*, 79, 777–781

Kuipers, E.J., Pérez-Pérez, G.I., Meuwissen, S.G. & Blaser, M.J. (1995) Helicobacter pylori and atrophic gastritis: importance of the cagA status. *J. Natl Cancer Inst.*, 87, 1777–1780

Kuo, G., Choo, Q.-L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., Tegtmeier, G.E., Bonino, F., Colombo, M., Lee, W.S., Kuo, L., Berger, K., Shuster, B.R., Overby, L.R., Bradley, D.W. & Houghton, M. (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science*, 244, 362–364

Lau, J.Y.N., Davis, G.L., Kniffen, J., Qian, K.P., Urdea, M.S., Chan, C.S., Mizokami, M., Neuwald, P.D. & Wilber, J.C. (1993) Significancer of serum hepatitis C virus RNA levels in chronic hepatitis C. *Lancet*, 341, 1501–1504

Lee, H.S., Han, C.J. & Kim, C.Y. (1993) Predominant etiologic association of hepatitis C Virus with hepatocellular carcinoma compared with hepatitis B virus in elderly patients in a hepatitis B-endemic area. *Cancer*, 72, 2564–2567

Levine, O.S., Vlahov, D. & Nelson, K.E. (1994) Epidemiology of hepatitis B virus infections among injecting drug users: seroprevalence, risk factors, and viral interactions. *Epidemiol. Rev.*, 16, 418–436

Liou, T.-C., Chang, T.-T., Young, K.-C., Lin, X.-Z., Lin, C.-Y. & Wu,H.-L. (1992) Detection of HCV RNA in saliva, urine, seminal fluid, and ascites. *J. Med. Virol.*, 37, 197–202

Lungu, O., Sun, X.W., Wright T.C., J.R., Ferenczy, A., Richart, R.M. & Silverstein, S. (1995) A polymerase chain reaction-enzyme-linked immunosorbent assay method for detecting human papillomavirus in cervical carcinomas and high-grade cervical cancer precursors. *Obstet. Gynecol.*, 8, 337–341

Maia, M., Takahashi, H., Adler, K., Garlick, R.K. & Wands, J.R. (1995) Development of a two-site immuno-PCR assay for hepatitis B surface antigen. *J. Virol. Methods*, 52, 273–286

McMahon, B.J., Rhoades, E.R., Heyward, W.L., Tower, E., Ritter, D., Lanier, A.P., Wainwright, R.B. & Helminiak, C. (1987) A comprehensive programme to reduce the incidence of hepatitis B virus infections and its sequelae in Alaskan natives. *Lancet*, 2, 1134–1136

McMahon, B.J., Alberts, S.R., Wainwright, R.B., Bulkow, L. & Lanier, A.P. (1990) Prospective study in 1400 hepatitis B surface antigen-positive Alaska Native carriers. *Arch. Intern. Med.*, 150, 1051–1054

McMahon, B.J., Wainwright, R.W. & Lanier, A.P. (1991) The Alaska native HCC screening program: a population-based screening program for hepatocellular carcinoma. In: Tabor, E., Di Bisceglie, A.M. & Purcell, R.H., eds, *Etiology, Pathology, and Treatment of Hepatocellular Carcinoma in North America* (Advances in Applied Biotechnology Series, Vol. 13). Houston, PPC–Gulf Publishing Company, pp. 231–241

Mattson, L., Weiland, O. & Glaumann, H. (1989) Chronic non-A, non-B hepatitis developed after transfusion, illicit self-injections or sporadically: outcome during long-term follow-up—a comparison. *Liver*, 9, 110–127

Maynard, J.E., Kane, M.A. & Hadler, S.C. (1989) Global control of hepatitis B through vaccination: role of hepatitis B vaccine in the expanded programme on

immunization. *Rev. Infect. Dis.*, 11 (suppl. 3), s574–s578

Muñoz, N., Bosch, F.X., de Sanjosé, S., Tafur, L., Izarzugaza I. Gili, M., Viladiu, P., Navarro, C., Martos, C., Ascunce, N., Gonzalez, L.C., Kaldor, J.M., Guerrero, E., Lörincz, A., Santamaria, M., Alonso de Ruiz, P., Aristizabal, N. & Shah, K. (1992) The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Colombia and Spain. *Int. J. Cancer*, 52, 743–749

Muñoz, N. & Pisani, P. (1994) Helicobacter pylori and gastric cancer. *Eur. J. Gastroenterol. Hepatol.*, 6, 1097–1103

Muñoz, N. & Bosch, F.X. (1996) Current views on the epidemiology of HPV and cervical cancer. Papillomavirus report. In: Lacey, C., ed., *Papillomavirus Reviews: Current Research on Papillomaviruses*. Leeds, Leeds University Press. pp. 227–237

Muratori, L., Giostra, F., Cataleta, M., Francesconi, R., Ballardini, G., Cassani, F., Lenzi, M. & Bianchi, F.B. (1994) Testing for hepatitis C virus sequences in peripheral blood mononuclear cells of patients with chronic hepatitis C in the absence of serum hepatitis C virus RNA. *Liver*, 14, 124–128

Nakajima, E., Thiers, V., Larzul, D., Thiollais, P. & Brechot, C.H. (1989) Hepatitis B virus (HBV) DNA as a markers of HBV infection. In: Bannasch, P., Keppler, D. & Weber, G., eds, *Liver Cell Carcinoma. Falk Symposium No. 51*. Dordrecht, Kluwer Academic Publishers. pp. 107–111

Nedjar, S., Mitchell, F. & Biswas, R. (1994) Simultaneous amplification and detection of specific hepatitis B virus and hepatitis C virus genomic sequences in serum samples. *J. Med. Virol.*, 42, 212–216

Ohto, H., Terazawa, S., Sasaki, N., Hino, K., Ishiwata, C., Kako, M., Ujiie, N., Endo, C., Matsui, A., Okamoto, H., & Mishiro, S. (1994) Transmission of hepatitis C virus from the mothers to infants. *New Engl. J. Med.*, 330, 744–750

Okamoto, H., Okada, S., Sugiyama, Y., Tanaka, T., Sugai, Y., Akahane, Y., Machida, A., Mishiro, S., Yoshizawa, H., Miyakawa Y. & Mayumi, M. (1990) Detection of hepatitis C virus RNA by a two-stage polymerase chain reaction with two pairs of primers deduced from the 5'-noncoding region. *Jpn J. Exp. Med.*, 60, 215–222

Okuno, H., Xie, Z., Lu, B., Qin, X., Takasu, M., Kano, H., Shiozaki, Y. & Inoue, K. (1994) A low prevalence of anti-hepatitis C virus antibody in patients with hepatocellular carcinoma in Guangxi Province, southern China. *Cancer*, 73, 58–62

Ozturk, M. (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*, 338, 1356–1359

Paterlini, P. & Brechot, C. (1994) Hepatitis B virus and primary liver cancer in hepatitis B surface antigen positive and negative patients. In: Brechot, C., ed., *Primary Liver Cancer Aetiological and Progression Factors*. CRC Press, pp 167–190

Pisani, P., Parkin, D.M., Muñoz, N. & Ferlay, J. (1997) Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiol. Biomarkers Prev.* (in press)

Poll, S., Driss, F., Carnot, F., Michel, M.-L., Berthelot, P. & Brechot, C. (1993) Vaccination against hepatitis B virus; an efficient immunotherapy against hepatitis B multiplication. *C.R. Acad Sci. Paris, Life/Sciences*, 316, 688–691

Qian, C., Camps, J., Maluenda, M.D., Civeira, M.P. & Prieto, J. (1992) Replication of hepatitis C virus in peripheral blood mononuclear cells. Effect of alpha-interferon therapy. *J. Hepatology*, 16, 380–383

Qian, G.-S., Ross, R.K., Yu, M.C., Yuan, J.-M., Gao, Y.-T., Henderson, B.E., Wogan, G.N. & Groopman, J.D. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol. Biomarkers Prev.*, 3, 3–10

Quint, W.G., Heijtink, R.A., Schirm, J., Gerlich, W.H. & Niesters, H.G. (1995) Reliability of methods for hepatitis B virus DNA detection. *J. Clin. Microbiol.*, 33, 225–228

Ranki, M., Schatzl, H.M., Zachoval, R., Uusi-Oukari, M. & Lehtovaara, P. (1995) Quantification of hepatitis B virus DNA over a wide range from serum for studying viral replicative activity in response to treatment and in recurrent infection. *Hepatology*, 21, 1492–1499

Ross, R.K., Yuan, J.-M, Yu, M.C., Wogan, G.N., Qian, G.-S., Tu, J.-T., Groopman, J.D., Gao, Y.-T. & Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, 339, 943–946

Sakuma, K., Saitoh, N., Kasai, M., Jitsukawa, H., Yoshino, I., Yamaguchi, M., Nobutomo, K., Yamumi, M., Tsuda, F., Komazawa, T.,Nakamura, T., Yoshida, Y. & Okud, K. (1988) Relative risks of death due to liver disease among Japanese male adults having various statuses for hepatitis Bs and e antigen/antibody in serum: a prospective study. *Hepatology*, 8, 1642–1646

Schiffman, M.H., Kiviat, N.B., Burk, R.D., Shah, K.V., Daniel, R.W., Lewis, R., Kupers, J., Manos, M.M., Scott, D.R., Sherman, M.E., Kurman, R.J., Stoler, M.H., Glass, A.G., Rush, B.B., Mielzynska, I. & Lörincz, A.T. (1995) Accuracy and interlaboratory reliability of human papillomavirus DNA testing using Hybrid Capture. *J. Clin. Microbiol.*, 33, 545–550

Schmid, M., Piroving, M., Altorfer, J., et al (1982) Acute hepatitis non-A, non-B: are there any specific light microscopic features? *Liver*, 2, 61–67

Seelig, R., Renz, M. & Seelig, H.P. (1992) PCR (polymerase chain reaction) in the diagnosis of viral hepatitis. *Ann. Med.*, 24, 225–230

Sherman, M.E., Schiffman, M.H., Lorincz, A.T., Manos, M.M., Scott, D.R., Kurman, R.J., Kiviat, N.B., Stoler, M.,

Glass, A.G. & Rush, B.B. (1994) Toward objective quality assurance in cervical cytopathology. Correlation of cytopathologic diagnoses with detection of high-risk human papillomavirus types. *Am. J. Clin. Pathol.*, 102, 182–187

Sherman, M., Peltekian, K.M. & Lee, C. (1995) Screening for hepatocellular carcinoma in chronic carriers of hepatitis B virus: incidence and prevalence of hepatocellular carcinoma in a North American urban population. *Hepatology*, 22, 432–438

Simmonds, P., Alberti, A., Alter, H.J., Bonino, F., Bradley, O.W. & Brechot, C. (1994) A proposed system for the nomenclature of hepatitis C virus genotypes. *Hepatology*, 19, 1321–1324

Stewart, A.-C., Eriksson, A.M., Manos, M.M., Muñoz, N. & Wheeler, C.M. (1996) Intratype variation in 12 human papillomavirus types: a worldwide perspective. *J. Virol.*, 70, 3127–3136

Stroffolini, T., Chiaramonte, M., Tiribelli, C., Villa, E., Simonetti, R.G., Rapicetta, M., Stazi, M.A., Bertin, T., Croce, S.L., Trande, P., Magliocco, A. & Chionne, P. (1992) Hepatitis C virus infection, HBsAg carrier state and hepatocellular carcinoma: relative risk and population attributable risk from a case-control study in Italy. *J. Hepatology*, 16, 360–363

Su, Q., Liu, Y., Zhang, S., Li, D. & Yang, J. (1994) Expression of insulin-like growth factor II in hepatitis B, cirrhosis and hepatocellular carcinoma: its relationship with hepatitis B virus antigen expression. *Hepatology*, 19, 788–799

Szmuness, W. (1978) Hepatocellular carcinoma and the hepatitis B virus: evidence for a causal association. *Prog. Med. Virol.*, 24, 40–69

Tabor, E. (1994) Tumor suppressor gener, growth factor genes, and ondogenes in hepatitis B virus-associated hepatocellular carcinoma. *J. Virol.*, 42, 357–365

Tanaka, K., Hirohata, T., Koga, S., Sugimachi, K., Kanematsu, T., Ohryohji, F., Nawata, H., Ishibashi, H., Maeda, Y., Kiyokawa, H., Tokunaga, K., Irita, Y., Takeshita, S., Arase, Y. & Nishino, N. (1991) Hepatitis C and hepatitis B in the etiology of hepatocellular carcinoma in the Japanese population. *Cancer Res.*, 51, 2842–2847

Thursz, M.R., Kwiatkowski, D., Allsopp, C.E.M., Greenwood, B.M., Thomas, H.C. & Hill, A.V.S. (1995) Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *New Engl. J. Med.*, 332, 1065–1069

Tito, L., Sánchez, J.M., Costa, J., Jove, J., Vilella, A., Bruguera, M. & Rodes, J. (1990) Long-term follow-up of chronic NANB (C) hepatitis C [abstract]. *J. Hepatology* (suppl. 2), S61

Tsen, Y.J., Chang, M.H., Hsu, H.Y., Lee, C.Y., Sung, J.L. & Chen, D.S. (1991) Seroprevalence of hepatitis B virus infection in children in Taipei, 1989: five years after a mass hepatitis B. vaccination program. *J. Med. Virol.*, 34, 96–99

Vall Mayans, M., Calvet, X., Bruix, J., Bruguera, M., Costa, J., Estève, J., Bosch, F.X., Bru, C. & Rodés, J. (1990) Risk factors for hepatocellular carcinoma in Catalonia, Spain. *Int. J. Cancer*, 46, 378–381

Villeneuve, J.P., Desrochers, M., Infante-Rivard, C., Willems, B., Raymond, G., Bourcier, M., Côté, J. & Richer, G. (1994) A long-term follow-up study of asymptomatic hepatitis B surface antigen-positive carriers in Montreal. *Gastroenterology*, 106, 1000–1005

Walboomers, J.M.M., de Roda Husman, A.-M., van den Brule, A.J.C., Snijders, P.J.F. & Meije, C.J.L.M. (1994) Detection of genital human papillomavirus infections: critial review of methods and prevalence studies in relation to cervical cancer. In: Stern, P.L. & Stanley, M.A., eds, *Human Papillomavirus and Cervical Cancer: Biology and Immunology*. Oxford, Oxford University Press, pp. 41–71

Whittle, H.C., Inskip, H., Hall, A.J., Mendy, M., Downes, R. & Hoare, S. (1991) Vaccination against hepatitis B and protection against chronic viral carriage in The Gambia. *Lancet*, 337, 747–750

Yuan, J.M., Ross, R.K, Stanczyk, F.Z., Govindarajan, S. & Gao, Y.T. (1995) A cohort study of serum testosterone and hepatocellular carcinoma in Shanghai, Chaina. *Int. J. Cancer*, 63, 419–493

Zaaijier, H.L., Cuypers, H.T.M., Reesink, H.W., Winkel, I.N., Gerken, G. & Lelie, P.N. (1993) Reliability of polymerase chain reaction for detection of hepatitis C virus. *Lancet*, 341, 722–724

Zhang, Y.Y., Hansson, B.G., Kuo, L.S., Widell, A. & Nordenfelt, E. (1993) Hepatitis B virus DNA in serum and liver is commonly found in Chinese patients with chronic liver disease despite the presence of antibodies to HBsAg. *Hepatology*, 17, 538–544

Corresponding author
N. Muñoz
Unit of Field and Intervention Studies, International Agency for Research on Cancer, 150 cours Albert Thomas, F-69372, Lyon Cedex 08, France

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Carcinogen–DNA and carcinogen–protein adducts in molecular epidemiology

C.P. Wild and P. Pisani

Carcinogen–DNA and carcinogen–protein adducts provide an integrated measure of carcinogen exposure, uptake and absorption, metabolism, DNA repair and cell turnover. As such they promise to provide a more objective and relevant measure of exposure than that which can be derived from questionnaires and measures of ambient levels of carcinogen. Nevertheless, the interpretation of adduct measurements made in human tissues and body fluids requires an understanding of a number of factors. These include the sensitivity and specificity of the measurement, the temporal relationship between exposure and adduct level and the mechanistic role of the adduct in the process of carcinogenesis. The application of such biomarkers in epidemiological studies therefore necessitates careful consideration of optimal study design. The above issues are illustrated in this chapter with examples from studies in both animal models and human populations.

Chemicals can covalently bind to cellular macromolecules including DNA, RNA and proteins. The product of this 'addition' of a chemical moiety to a macromolecule is termed an adduct. These adducts may be measured in tissues, exfoliated cells, peripheral blood or urine. Other types of DNA damage result from exposure to chemicals or through biological processes such as inflammation, e.g. 8-oxo-deoxyguanosine (8-oxo-dG), although these are not strictly adducts by the above definition. However, for simplicity these types of DNA damage are also referred to as 'adducts' in this paper. The measurement of carcinogen–nucleic acid or carcinogen–protein adducts is a promising approach to determining exposure to DNA-damaging carcinogens in humans. The degree to which such measurements can be quantitatively associated with cancer risk at the individual level is not understood.

The measurement of adducts is of potential value in exposure assessment for a number of reasons. First, it may provide a more objective and relevant measure of individual exposure than can be obtained by a questionnaire approach. Second, it is potentially highly specific for the exposure of interest. The specificity of adducts makes them particularly suitable when defined compounds/carcinogens are investigated. Third, the measurements are often of a high sensitivity. There are, however, numerous issues that can impinge on the validity of this type of measurement. These issues are of both a conceptual and a practical nature. This paper therefore aims to evaluate current knowledge regarding the validity of adducts as measures of carcinogen exposure and cancer risk in different epidemiological study designs.

Rationale for use of adducts in exposure assessment

The rationale for using measurements of carcinogen–DNA and carcinogen–protein adducts in human exposure assessment is based on the assumption that DNA adducts formed *in vivo* are responsible for genetic alterations in genes critical for carcinogenesis and that the protein adducts formed through the same processes reflect the formation of DNA adducts. From this basic assumption, a number of extrapolations are made to allow the measurements to be carried out with the available methodologies in available human samples, e.g. in surrogate tissues as opposed to the target tissue. As presented in Fig. 1, these extrapolations include that from the specific DNA sequence level within specific genes to the level of genomic DNA, that from the single cell to the target tissue or organ, and that from the target organ to peripheral blood. In addition, as mentioned above, extrapolations are also made from DNA to protein adducts. This illustrates that, in the majority of instances, what

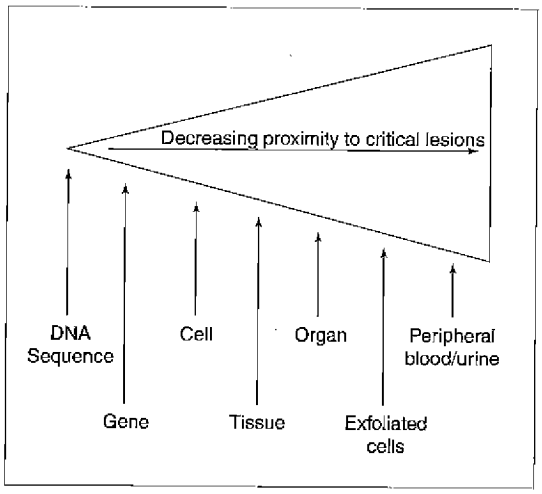

Figure 1. Measurement of carcinogen–DNA and DNA–protein adducts: proximity to critical lesion.

is actually measured is far removed from what one would ideally measure. Such 'surrogate' measures require validation for their relationship to the 'ideal' measure.

An example of the above issue is the case of aflatoxins where a specific mutation at the third nucleotide of codon 249 of the *p53* tumour suppressor gene in human hepatocellular carcinomas (HCCs) has been suggested to result from aflatoxin exposure (Ozturk, 1991). Assuming, for argument's sake, that this is the case, the 'ideal' measure would be aflatoxin–DNA adducts in codon 249 of the *p53* gene in human hepatocytes. In reality, a few measurements have been performed on total genomic DNA from liver (Zhang *et al.*, 1991) but most epidemiological studies have been performed using aflatoxin–nucleic acid adducts excreted in the urine (Autrup *et al.*, 1987; Groopman *et al.*, 1993) or aflatoxin–albumin adducts in serum (Wild *et al.*, 1993). Nevertheless, the presence of urinary AFB1-*N*7-guanine adduct was associated with an increased risk of HCC in a prospective cohort study in the People's Republic of China (Qian *et al.*, 1994). It is noteworthy that, prior to its successful application in this study, this biomarker had undergone extensive validation (Groopman, 1994) to establish the dose–response relationships between exposure and adduct levels in liver, urine and albumin, both in animal models and in human populations exposed to dietary aflatoxins.

Determinants of adduct levels *in vivo*

Once a chemical carcinogen enters the body, the level of adduct formation will depend upon a number of factors, including the absorption and distribution of the chemical around the body and the organ-specific and cell-specific metabolism (activation and detoxification). The measured adduct level at any one point in time will further depend on the chemical stability of the adduct, any repair processes (for DNA adducts) and cell turnover. Thus, the carcinogen–DNA adduct represents an integration of these different parameters at the individual level, and measures the dose of carcinogen reaching the target molecule for carcinogenesis; this has been termed the 'biologically effective dose'. It is therefore an integration of both exposure and interindividual variations in metabolism and DNA repair, each of which may be determined genetically and/or be influenced by other environmental exposures. Some DNA adducts are removed by excision repair processes, e.g. *N*-methylpurine glycosylases (Bessho *et al.*, 1993), or are lost spontaneously, as for example in depurination, and these adducts can be detected in the urine as free bases or deoxynucleosides (Shuker & Farmer, 1992). The carcinogen–protein adducts, while not representing events thought to be critical in carcinogenesis, are of interest because they may reflect the level of DNA adducts in internal organs (Skipper & Tannenbaum, 1990; Skipper *et al.*, 1994).

Dose–response studies: adducts, mutations and cancer

The above rationale for adduct measurement depends on the proposed causal link between DNA adducts, mutations and tumour induction. DNA adducts induced by specific carcinogens have been linked to mutations in a number of different types of study, including shuttle vectors containing adducts of specific carcinogens (e.g. Maher *et al.*, 1989; Kat & Thilly, 1994). Mutation spectra in animal tumours in genes involved in the carcinogenic process are generally consistent with those expected from a knowledge of the types of DNA adduct induced by specific chemicals (Greenblatt, 1994).

Poirier and Beland (1992) reviewed data in experimental animals where steady-state DNA adduct levels after 1–2 months' chronic exposure have been compared with tumour induction following lifetime exposure to the carcinogen, e.g.

aflatoxin B_1 (AFB1) for rat and trout liver; diethylnitrosamine for rat liver; 2-acetylaminofluorene (2-AAF) for mouse liver and bladder; ethylene oxide for rat spleen mononuclear cell leukemia; 4-methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) for rat lung tumours; and 4-aminobiphenyl (4-ABP) for mouse liver and bladder tumours. For some of these chemicals, a linear dose–response for both DNA adducts in the target organ and tumour induction were observed (e.g. rat and trout liver with AFB1). However, in other situations, including bladder of female mice with 2-AAF and male mice with 4-ABP, adduct levels increased at all doses but tumours were only induced at the highest doses (Poirier & Beland, 1992; Poirier *et al.*, 1995). In the case of the 2-AAF experiments, the tumour induction was seen at dose levels associated with increased cell proliferation in the bladder (Cohen & Ellwein, 1990).

There are examples of differences in response to carcinogens between males and females apparently as a result of differences in cell proliferative response rather than DNA adduct formation. One example is that of rats treated with 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) (Ochiai *et al.*, 1996). In this experiment rats were given PhIP in the diet for up to 12 weeks, and DNA adducts, cell proliferation and aberrant crypt foci were measured in the colon. PhIP–DNA adduct levels were the same in male and female rats, but the incidence of aberrant crypt foci was threefold higher in male than in female rats. This increase was associated with a 42% increase in labelling index in the male colon compared with the female colon after 8 weeks' treatment.

In the studies of rat lung with NNK, the determination of DNA adducts in specific cell populations within the lung was important if the dose–response curve for tumour induction in that organ was to be understood (Belinsky *et al.*, 1987, 1990). This study illustrated that adduct measurements in DNA extracted from total tissue are not necessarily related to tumour induction, but rather that the target cell population needs to be identified.

The studies cited above compared the DNA adduct levels in one set of animals with the tumour incidence in a parallel set of animals. The former were treated with carcinogens under the same regimen but for a shorter duration, sufficient to allow steady-state adduct levels to be reached. There are few studies which have related DNA adducts and tumour incidence at the level of the individual animal. One attempt to perform such an experiment was reported by Fischer and Lutz (1995) who examined the induction of papillomas on skin after exposure of outbred mice to 7,12-dimethylbenz[*a*]-anthracene (DMBA). These authors used the latency period to appearance of the first papilloma as an indicator of individual risk. DMBA–DNA adducts, 8-OHdG adducts and cell division were determined in the dermis of the treated skin area 2 weeks after the appearance of the first papilloma. The data indicated a positive correlation between shorter latency period of papilloma induction and both 8-OHdG levels and cell proliferation. However, DMBA–DNA adducts were significantly higher in the animals with longer latency period. This was explained by the lower rate of cell division in these animals compared with those with short latency, resulting in a slower 'dilution' of the DNA adduct concentration after the end of treatment. This again indicates that measurement of carcinogen–DNA adducts does not always reflect individual risk of tumour induction. The data on 8-OHdG are of particular interest given the growing body of evidence showing that secondary reactions resulting from carcinogen exposure could be of importance in addition to the primary formation of DNA adducts (see below).

These above data suggest that DNA adducts are necessary but not sufficient to result in tumours with these chemical carcinogens. Consequently, while measurements of DNA adducts in a given tissue are certainly evidence of exposure of the target organ to the carcinogen, they are unlikely to be sufficiently informative alone to predict tumour incidence in that organ at the individual level. The important role of cell proliferation both in carcinogenesis and in determining adduct levels is supported by the above observations. The degree to which the above mechanisms interfere with the association between exposure and DNA adduct levels in human populations needs to be taken into account when incorporating these markers into epidemiological studies.

Methods for measuring carcinogen adducts

Many sophisticated techniques have been developed to measure adducts in human tissues and cells. These generally include a clean-up step(s) followed by an analytical step. Clean-up steps comprise solvent

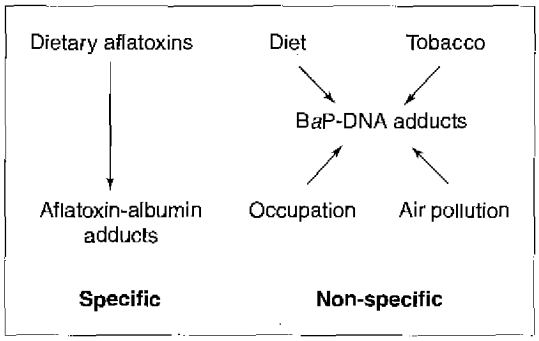

Figure 2. Is the adduct specific for the exposure under study? BaP, benzo[*a*]pyrene.

extractions, solid-phase extraction, immunoaffinity, liquid and gas chromatography, etc. Analytical techniques comprise mass spectrometry, immunoassays, fluorescence, UV or electrochemical detection, etc. These approaches all have inherent strengths and weaknesses which have been reviewed in depth in a number of comprehensive reviews and are not repeated here (Wogan, 1988, 1992; Shuker & Farmer, 1992). In general, the above methods are of a high sensitivity, i.e. they can measure the adduct at low levels in the biological sample. For example, the ^{32}P-postlabelling technique has permitted detection of adducts at the level of one adduct per mammalian genome (Chacko & Gupta, 1988). The value of such sensitivity is illustrated by the ability to identify 4-aminobiphenyl–Hb adducts in the blood of non-smokers exposed to environmental tobacco smoke (Hammond *et al.*, 1993). The presence of carcinogen adducts in human cells and body fluids following environmental, occupational and therapeutic exposures has been documented and reviewed (Wogan, 1992).

Interlaboratory comparisons of methods have been relatively few and appear to have been organized mainly on an *ad hoc* basis between research groups. An exception to this is an international collaborative study organized by the US Environmental Protection Agency and supported by the European Union, in collaboration with the NCI and NCTR, USA, coordinated by the IARC, France (Dr M. Castegnaro), and the CRC, UK (Dr D. Phillips). This collaboration aims to develop standard protocols for the analysis of specific adducts by the ^{32}P-postlabelling approach. In addition, DNA adduct standards for different chemical carcinogens are being prepared. This type of method standardization is clearly an important process if large-scale, multicentre studies are to be performed with sample analysis being conducted in different laboratories.

Multiple adducts from individual carcinogens

As mentioned above, the specificity of adduct measurements for exposure to a particular carcinogen is one of the potential benefits of this approach. Measurement of specific adducts is based on the elucidation of the chemical structure of the adduct. With this information, adduct standards can be prepared, antibodies raised to the adduct, the fluorescent, UV and electrochemical properties etc. investigated, and appropriate analytical techniques developed. It is this basis that permits the measures to be specific for the agent of interest. Other methodologies to measure DNA adducts, notably the ^{32}P-post-labelling technique, as originally developed, or immunoassays analysing intact DNA, forfeit a degree of specificity but are appropriate as an approach to measuring exposure to a defined class of carcinogens. This is valuable in situations of exposure to complex mixtures of chemicals (Perera *et al.*, 1990).

To date, several hundred carcinogen–DNA adducts have been identified (Hemminki *et al.*, 1994). In the majority of cases even a specific carcinogen forms adducts at different sites on the same molecule. For example, a simple environmental methylating agent such as dimethyl-nitrosamine results in at least 12 different DNA adducts (Margison & O'Conner, 1979). In this case the most abundant adduct, 7-methylguanine, does not appear to be the most relevant for induction of mutations but rather the minor O-methylated bases, O^6-methylguanine (O^6-meG) and O^4-methylthymidine, are the promutagenic adducts (Pegg, 1984; Richardson *et al.*, 1987).

A compound may also form chemically distinct types of adduct, which differ in their degree of specificity for a given exposure. For example, the tobacco-specific nitrosamine (TSNA) NNK can induce both methylation adducts and pyridyl-oxobutylation adducts. While the O-methylation adducts (e.g. O^6-meG) may be the source of mutations induced by this compound, the adduct is not specific to TSNA but is also formed as a result of exposure to other environmental methylating agents as well as to endogenously formed methylating agents (Bartsch & Montesano, 1984). In this case, the pyridy-

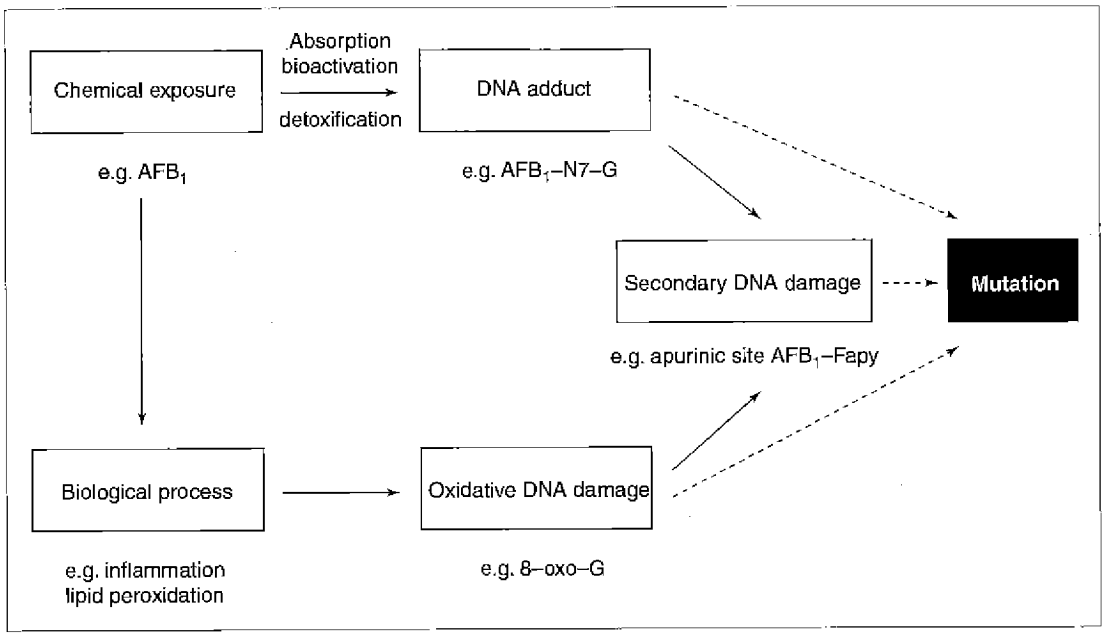

Figure 3. Formation of DNA damage via different pathways.

loxobutyl adducts would be a more specific marker for exposure to TSNA.

A highly specific measure of a given adduct is therefore not equivalent to having a highly specific marker for a given exposure. For example, while aflatoxin–albumin adducts are specific for dietary exposure to aflatoxins, benzo[*a*]pyrene–DNA adducts can result from tobacco smoke, air pollution, diet and/or occupation (Fig. 2). In this case, use of this latter marker as a measure of exposure to air pollutants would need to be complemented by measures of the possible confounding exposures mentioned.

It is necessary to ensure that adducts do not result from sources other than from the covalent binding of a chemical to a macromolecule in the body. Specific examples are the difficulties of discriminating the naturally occurring ribonucleoside 7-methylguanosine from the DNA adduct 7 methyldeoxyguanosine (Bianchini *et al.*, 1993) and the presence of 3-methyladenine as a natural component of the diet (Shuker & Farmer, 1992).

A further complication regarding specificity and relevance of adduct measurements is that the initial adduct formed in DNA may not be the critical lesion involved in mutagenesis. This is illustrated by the example of AFB1 presented in Fig. 3. This carcinogen is activated to an AFB1 8,9-epoxide which binds predominantly, although perhaps not exclusively, to the *N*7 position of guanine (Martin & Garner, 1977; Essigmann *et al.*, 1977). However, this adduct is highly unstable and through either active DNA repair or spontaneous depurination results in formation of an apurinic site in DNA. The mutation spectra observed in bacterial systems, in mammalian cells *in vitro* and in rodent and human tumours are consistent with a miscoding event resulting from this apurinic site (Trottier *et al.*, 1992). The free AFB1-*N*7-guanine base is excreted in the urine and can be detected by immunoaffinity/HPLC as a marker of human exposure (Groopman *et al.*, 1993); this is a good marker of initial hepatic DNA adduct levels (Groopman *et al.*, 1992). However, the AFB1-*N*7-guanine in DNA can also undergo a chemically induced imidazole-ring cleavage such that the major DNA adduct in rat liver 48 h post-treatment is the imidazole ring-open AFB1-*N*7-guanine, known as AFB1-Fapy (Hertzog *et al.*, 1982). Therefore, the adducts measured in human liver tissue are most probably the AFB1-Fapy rather than the presumed mutagenic lesion. The situation is further complicated by the fact that, at least in rat

liver, treatment with AFB1 results not only in direct binding of AFB1 to DNA but also in an oxidative stress with a consequent increase in oxidative DNA damage (Shen *et al.*, 1994, 1995). In this type of exposure situation, the choice of the most informative biomarker would be facilitated by an understanding of which adduct is involved in the generation of mutations in critical genes for carcinogenesis. The AFB1-*N7*-guanine adduct and the AFB1–albumin adduct are the most specific markers of exposure to AFB1, but the complementary information from measurements of oxidative damage in the same subjects would be valuable. It is interesting in the above example that both AFB1 and oxidative stress appear to preferentially target mutations to the third nucleotide of codon 249 of the *p53* gene (Aguilar *et al.*, 1993; Hussain *et al.*, 1994), a mutation in human hepatocellular carcinomas that has been suggested to be induced by aflatoxin exposure (Greenblatt *et al.*, 1994).

The need to understand the adducts involved in mutagenesis is further illustrated by a study of mutational spectra in the c-Ha-*ras* gene in mouse skin papillomas induced by several aromatic hydrocarbons (Chakravarti *et al.*, 1995). In this study the authors argued that the mutations (G to T or A to T transversions) were consistent with apurinic sites resulting from depurination at guanines and adenines rather than being the result of stable DNA adducts, the latter being what would normally be measured in molecular epidemiology studies. For example, this observation was suggested to be the case with benzo[*a*]pyrene which forms a significant proportion of stable C-8 guanine adducts, which have been the basis of many measurements of carcinogen exposure in human tissues. The above data at least suggest that assays of apurinic sites would be a valuable addition to methods for monitoring DNA damage in humans.

Overall, these examples illustrate the need for detailed information on adduct formation and the mutagenic potential of those adducts in choosing an appropriate end-point for exposure assessment. In addition, the need to understand the different exposures that can yield a specific chemical carcinogen adduct is indicated.

Protein adducts as alternatives to DNA adducts

Although proteins are not the target for mutational events, they are nevertheless an attractive alternative to DNA as target molecules (Skipper & Tannenbaum, 1990; Skipper *et al.*, 1994). This is because proteins are often available in larger amounts than DNA and also because the lack of repair of protein adducts results in a greater persistence (see below), and therefore the measured level is more closely correlated with past exposure. In practice, in epidemiological studies the measurement of protein adducts has been restricted to albumin (Alb) and haemoglobin (Hb) because of the ready availability of these proteins in peripheral blood. The identification of adducts on histones has been reported for some carcinogens (Ozbal *et al.*, 1994). The longer half-life of histones compared with Alb and Hb is one advantage of their use as dosimeters, as is the proximity of histones to DNA. In addition, histones may also play a role in gene regulation, and, consequently, the modification of histones by carcinogens may have a functional role in carcinogenesis. The use of other long-lived proteins, including collagen, has also been discussed and could be applicable to molecular epidemiological studies in the future (Skipper *et al.*, 1994).

Target organs and surrogate tissues

As mentioned above, one of the potential advantages of examining DNA adducts is that they can be measured in the target cells for carcinogenesis. However, in many cases these cells are not available and alternative sources of DNA or proteins are used. Many studies have utilized DNA from peripheral blood cells (PBC), bronchiolar lavages or oral cavity, as well as adducts excreted in the urine. In only a few cases has the association between adduct levels in the target organ and those in the surrogate tissue or body fluid been examined in humans.

In animal studies, the ratio between DNA adducts in liver and PBC has been examined for a number of carcinogens and these data were recently reviewed (Bianchini & Wild, 1994a, 1994b). The ratios varied from 0.1 to 100 with even the small number of carcinogens studied. In a series of four methylating agents having different target organs for carcinogenesis, the ratio between liver and PBC for 7-meG levels was relatively constant, but the ratio between the target organ (lung, colon, oesophagus and liver) and PBC ranged from 3.6 to 200. These data suggest that the methylation of PBC DNA is occurring in the liver by transfer of

active methylating species from the hepatocyte to the PBC.

In humans, aromatic DNA adducts in buccal mucosal cells were highly correlated with levels of the same adducts in oral biopsies from the same individuals (Stone *et al.*, 1995). This type of correlation might be expected in cells coming from the same tissue, especially where cell turnover is rapid and therefore adduct levels reflect mainly recent exposure. There was also a good correlation between 7-meG levels in bronchial tissue and PBC in one study based on only five subjects. A low but significant correlation of 0.34 between polycyclic aromatic hydrocarbon (PAH)–DNA adducts in PBC and lung cancer tissue was reported in the study by Tang *et al.* (1995). In the same study, the DNA adduct levels in PBC did not correlate with levels in non-tumour lung tissue. Another study reported no correlation between the total PAH–DNA adduct levels in lung and PBC, but a significant correlation in the case of the major DNA adduct (van Schooten *et al.*, 1992). Aromatic DNA adducts measured by ^{32}P-labelling assay were highly correlated in non-tumour lung tissue of lung cancer patients and peripheral blood mononuclear cells ($r = 0.74$) (Wiencke *et al.*, 1995); moreover, adduct levels in both tissues correlated negatively with years since quitting smoking. The study suggested that aromatic DNA adducts in peripheral blood mononuclear cells could be an indicator of the DNA damage caused by smoking in lung tissue. Phillips *et al.* (1990) carried out a similar study, looking at aromatic DNA adducts in non-affected lung tissue and PBC. They found that adducts in the lung correlated with smoking habits, but that the marker assessed in PBC did not. In around 10 larynx cancer patients, the levels of aromatic DNA adducts as determined by ^{32}P-postlabelling in larynx tumour tissue and surrounding non-tumorous tissue correlated well with adducts in PBC (Szyfter *et al.*, 1994). Thus, to date, there are some published data to support a correlation between DNA adducts in PBC and those in internal organs, although the limited number of studies should be noted and further similar studies conducted where possible.

Persistence of adducts

Once an adduct has been formed, its persistence will depend on a number of factors, namely its inherent chemical stability, whether any active repair processes are present, and the turnover of the macromolecule to which the chemical is bound. This parameter of persistence is particularly critical in determining the appropriate epidemiological study design in which to apply adduct measurements.

In the case of DNA adducts, the chemical stability is highly variable. For example, the AFB1-*N*7-G adduct is lost by depurination with a half-life of around 8 h in rat liver (Groopman *et al.*, 1992). A proportion of the initial adducts are, however, susceptible to imidazole ring-opening to form a more stable AFB1-Fapy adduct which has been reported to persist in rat liver DNA as long as 19 weeks after the last treatment.

In addition to the inherent chemical stability of the DNA adduct, there are also active repair enzymes. For example, in human cells, the methylated DNA bases *N*7-methyldeoxyguanosine and *N*3-methyldeoxyadenosine, as well as 8-oxo-guanine, are repaired by the *N*-methylpurine glycosylase enzyme, although at markedly different rates (Male *et al.*, 1987; Bessho *et al.*, 1993). The O^6-meG adduct is repaired efficiently in most human tissues by the O^6-alkylguanine DNA alkyltransferase. The situation is rendered more complex by the fact that DNA repair activity is not uniform throughout the genome, with more rapid repair of some types of DNA adduct occurring in transcribed as opposed to non-transcribed genes and with a strand bias for the transcribed strand (Bohr, 1991; May *et al.*, 1993). This appears to be of importance in the process of carcinogenesis as there is a predominance of mutations in the non-transcribed strand of the *p53* tumour suppressor gene (Greenblatt *et al.*, 1994).

Persistence of adducts is also dependent on the stability of the molecule to which the chemical is bound. In the case of DNA this is essentially translated as the rate of cell turnover. Thus adducts in epithelial cells lining the large bowel or in neutrophils in peripheral blood (half-life ≈ 8 h) will be far less persistent than adducts in non-proliferating tissues such as brain or liver or in long-lived T-lymphocytes. In fact, one would predict that DNA adduct levels in the same individual should be higher in lymphocytes than total leukocytes. This has been rarely studied, but in those cases where it has been examined the levels were higher in the longer-lived cells (Mustonen & Hemminki, 1992). It is interesting that the differ-

ences in adduct levels between smokers and non-smokers in this study were observed in total leukocytes, granulocytes and lymphocytes, suggesting that recent exposure over the past 1 or 2 days was correlated with past exposure of a few months. The differences may not be so consistent in exposures of more intermittent pattern.

One consequence of the above complexity of DNA adduct formation and persistence is that half-lives of different DNA adducts need to be empirically determined. A second consequence is that most measurements of DNA adducts in human organs will be an average measure of events at the sequence, gene, cell and tissue levels.

The situation for carcinogen–protein adducts is somewhat simpler than for the DNA adducts. In general it has been assumed that in the absence of active repair the adducted protein would have the same half-life as the unadducted molecule—i.e. in the case of haemoglobin, a half-life of around 120 days, and for albumin adducts around 20 days. Surprisingly, this has not been confirmed for many adducts by longitudinal studies either in animals or humans. Carmella and Hecht (1987) examined the half-life of a tobacco-specific nitrosamine–Hb adduct in rats and reported that, with a half-life of 9.1 days, this was significantly shorter than expected. Given the first-order kinetics of adduct loss, this was most probably the result of instability of the adduct rather than a preferential clearance of the chemically modified haemoglobin.

In humans, only a few studies have been reported, for the most part relating to subjects who stopped smoking for an agreed period of time. For example, in a study of 4-aminobiphenyl–Hb (4-ABP–Hb) adducts, smokers stopped smoking for a period of up to 80 days and the observed rate of decline of the adducts was significantly in excess of that expected (Maclure *et al.*, 1990). In the study by Mooney *et al.* (1995) both 4-ABP–Hb and PAH–DNA adduct levels were monitored in volunteers for up to 14 months after they stopped smoking. Both markers were significantly decreased 10 weeks after smoking cessation and the half-lives were estimated to be 23 weeks (95% CI, 10.5–36.3) and 12 weeks (95% CI, 9.9–14.0) for PAH–DNA adducts and 4-ABP–Hb adducts respectively.

Adduct stability during storage and analysis

The attraction of applying measures of adducts in prospective studies is easy to appreciate. However, a particular issue of concern in the context of long-term storage of biological samples is the stability of adducts. This is of concern both in the initial phases of sample collection and processing and in the storage of biological samples in freezers over many years. Unfortunately there are few data on this topic. It is generally accepted that rapid processing of biological samples to storage at low temperatures is desirable to avoid continued enzyme activity, which would degrade DNA and proteins or repair adducts after isolation. Postmortem stability of DNA has been examined (Bär *et al.*, 1988). However, there is little information on the impact on adduct levels in, for example, peripheral blood cells that have been isolated and frozen immediately after isolation or several hours later. Similarly, there are few data on the levels of adducts in the same samples after storage under different temperatures for different lengths of time. The stability of benzo[*a*]pyrene diolepoxide adducts in rat liver, lung and heart has been tested under various conditions (Izzotti *et al.*, 1993). These authors reported that adducts were stable in tissues up to 72 h at 4°C, but significant decreases were observed at 20°C or 37°C even after periods as short as 24 h. Under conditions that mimicked autopsy conditions, i.e. 16 h at 20°C or 16 h at 20°C followed by 24 h at 4°C, no significant variation in adducts was observed. Similar studies for other adducts, in different tissues and under different conditions of storage, would be of value.

Another issue is the artefactual formation of adducts during sample processing, where the reaction that leads to adduct formation *in vivo* can occur during sample storage or analysis. This is possibly best illustrated by the situation with oxidized bases where oxidation reactions must be avoided during sample preparation and analysis. This can occur during DNA preparation (Claycamp, 1992) or during analytical procedures, e.g. the formation of 8-oxo-dG and other bases during the silylation process prior to GC-EIMS analysis (Douki *et al.*, 1996). The appearance of hydroxyethylvaline in haemoglobin during storage of globin under different conditions has also been noted (Tornqvist, 1990). While the artefactual formation of adducts is not likely to be a problem with chemical car-

cinogens, it is worth noting that it is not only limited to oxidized bases. Scates *et al.* (1995) recently illustrated the artefactual formation of DNA adducts by bile acids *in vitro*, and the danger of induction of UV photoadducts during DNA purification has been noted (Widlak *et al.*, 1995).

The possible artefactual formation of DNA adducts therefore merits attention during method development.

Epidemiological study design and adduct measurements

A number of the above aspects of carcinogen–DNA and carcinogen–protein adducts have implications for the appropriate use of adducts in epidemiological studies. These are now discussed in relation to specific study designs. A comprehensive review on this topic has been published by Rothman *et al.*, (1995).

Application of adducts in case–control studies is limited when current and past exposures do not correlate within individuals; this may happen as a direct consequence of the disease, e.g. dietary habits, or because the exposure is not related to personal habits and, therefore, is more likely to change over time, e.g. occupational activities. In the case of diet and occupation, objective markers of exposure to specific chemicals would greatly increase the informativeness of the studies. However, to date, adduct measurements have not generally offered an improvement over questionnaire data on exposure in case–control studies. In a case–control study of aflatoxin exposure and liver cancer the aflatoxin–albumin adduct level was correlated with measurements of aflatoxins in household foods in the controls but not in the cases (Hall & Wild, 1994). This suggests that either the dietary intake of aflatoxins was affected by the disease or the disease itself modified the metabolism of the carcinogen (Hall & Wild, 1992). In a study of lung cancer, PAH–DNA adducts in PBC of cases were significantly higher than in controls, with an odds ratio of 7.7 (95% CI, 1.7–34, $P<0.01$) (Tang *et al.*, 1995). In this study, adduct levels were significantly associated with questionnaire-derived indices of smoking among the cases but not among the controls, suggesting a difference in biological response to smoking between these two groups.

In case–control studies of occupational cancer, retrospective assessment of exposure to specific chemicals in the workplace can be performed by linking job descriptions obtained normally by interview with documented usage and presence of the chemicals in the occupational activity or industrial process described. This linkage can be systematized in the so-called job-exposure matrices (JEMs) (Coggon *et al.*, 1984). The JEM approach represented a substantial improvement on the classical analysis of cancer risks by job title, because JEMs provide a reference category of non-exposed individuals and increase the study power by pooling all the individuals exposed to the same agent from different jobs into one 'exposed' group. However, there are limitations in that categories of exposure are only probabilistic; intensity of exposure cannot be assessed on a common scale for the majority of job titles in the population; high misclassification rates across exposure categories have been documented; and, finally, transferability of JEMs developed for one specific study to a new investigation is limited (Goldberg *et al.*, 1993).

In spite of these limitations of JEMs, cross-sectional individual assessment of internal dose or biologically effective dose (adducts) cannot replace questionnaire information, because current exposure is a bad indicator of past exposure in this instance. Most occupation-related cancer cases occur in retirement age, and even in the active age groups the proportion of those who move from one economic activity to another, changing environment, work tasks and consequently exposures, tends to be rather high. A better application of adduct measurements in this context is the validation of associations between jobs and cancer risk detected by case–control JEM-based studies, by determining average biological effective dose in workers classified as 'exposed' and 'non-exposed' by the matrix. This type of transitional study can substantially improve the evidence provided by a classical case–control study, and can also add specific knowledge to help in understanding the carcinogenic process.

Adducts as markers of exposure and effective dose have a good application in ecological studies, which relate exposure to a chemical experienced by different populations to their risk of the disease at the group level. In this case the disease does not affect the exposure level assessed in spite of the cross-sectional nature of the observation. Adduct markers have an ideal application when investi-

gating the effects of food-borne chemicals and air pollutants that cannot be measured at the individual level through interview. One example is the study of the relation between aflatoxin intake and the risk of hepatocellular carcinoma. Many ecological studies have been conducted on this subject, with aflatoxin intake indirectly estimated by the level of contamination of food samples. The results of these studies are heterogeneous and sometimes inconsistent. In one study of this type, conducted in Thailand, aflatoxin–albumin adducts were determined in population samples (Srivatanakul *et al.*, 1991). Another much large study in the People's Republic of China used urinary aflatoxin metabolites as a measure of exposure and found no correlation with liver cancer rates (Campbell *et al.*, 1990). In this case some of the urinary aflatoxin metabolites, in contrast to the AFB1-*N*7-G adduct, were later shown to be unrelated to dose and thus may have led to misclassification of exposure (Groopman, 1994). Ecological studies of the effect of air pollution in the general population could benefit from the assessment of adduct levels to a chemical commonly contaminating the atmosphere of urban areas. Many specific and non-specific chemicals and agents contaminate the air of urban areas, and this multiplicity of exposures may be seen as a serious limitation to the use of markers of specific exposures. However, it has been documented that urban air pollutants follow the same circadian, weekly and seasonal variations, which are mainly determined by atmospheric conditions and cyclic intensity of vehicle traffic. Therefore, a few sentinel markers may be good indicators of a general non-specific exposure to polluted urban air in addition to being markers of internal biologically effective dose to defined carcinogens. It should be mentioned, however, that ecological studies of air pollution should not rely only on markers such as adducts, which are the result of exposure to specific chemicals irrespective of the external source of that exposure. Consequently, while adducts can be incorporated into descriptive studies of this type to improve understanding and interpretation, it is necessary to carefully consider possible independent sources of exposure to the same chemicals as potential confounders. For example, a study comparing PAH–DNA adducts in a highly industrialized area of Poland with those in rural areas showed an increased PAH–DNA adduct level in peripheral blood cells of subjects from the former population (Perera *et al.*, 1992). The authors showed that after adjusting for current smoking, another potential source of PAH–DNA adducts, adduct levels were still positively associated with air pollution.

Among all observational studies, the prospective study is the design in which the potential for comparison biases is the least serious: the adduct is detected in material collected when the disease was not clinically manifest and, although the presence of latent disease cannot be excluded, analysis by time since exposure assessment can help in interpreting the results.

The most cost-effective way to analyse cohort studies, particularly when assessment of biological material is involved, is the nested case–control design, which consists of the analysis of the cases detected during the follow-up of the cohort, and of only an appropriate sample of those who do not develop the disease; this sample is in general a very small fraction of the entire cohort. The first example of a measure of a DNA adduct being associated with an excess cancer risk was provided using this type of study design (Ross *et al.*, 1992; Qian *et al.*, 1994). In this instance, AFB1-*N*7-G adducts in a single urine sample collected from a cohort in Shanghai, People's Republic of China, was associated with an increased risk of hepatocellular carcinoma. This type of study is needed to strengthen the rationale for using adducts as measures of outcome in epidemiological studies.

Besides potential confounders, which depend on the disease investigated, some additional matching of variables in nested case–control studies needs to be considered; these are the determinants of the accuracy of the adduct measurements previously described. Samples of cases and matched controls should be analysed in the same batches in order to distribute equally the effects of sample processing; in addition, controls should be selected so that their samples have been stored as long as those of their matched cases. This design avoids the introduction of comparison biases if preservation of the biological material affects the stability of adducts with time (see above). Long-term effects of preservation can be monitored within the study bank by storing control material, which is assayed at the time of storage to determine the baseline level and then re-assayed,

together with the material of cases and controls, each time that analysis is performed. However, it is difficult to ensure that assays performed years apart will be characterized by the same measurement error, and therefore long-term monitoring should aim at detecting only substantial effects of preservation.

One major advantage of prospective observations in the assessment of biological markers is that the collection of the biological material that will eventually be used to measure exposure can be scheduled in order to reduce the effect of intra-individual variation, particularly if it is of a cyclical nature. Examples of this are seasonal variation of grain contamination with aflatoxins and circadian variation of exposure to air pollutants in urban areas. These sources of random error can be controlled in a prospective design in two ways:

- by imposing restrictions on the scheduled time of collection at the enrolment stage of the cohort (more appropriate for the air pollution example);
- by matching nested controls to the cases for similar conditions at the time of enrolment, provided that the relevant information was recorded and is available.

Intra-individual variability that cannot be attributed to known and controllable external factors can instead be reduced with repeated samples from each individual. The timing and frequency of repeated measures are often governed by financial constraints rather than scientific considerations. However, the optimal use of available resources should be driven by some knowledge of the extent of the variation and its possible cyclic nature. Intra-individual variation must be separated from systematic time trends, e.g. hormonal levels in women vary in a regular fashion over days and months; at the same time, important trends are documented with age. As already mentioned, cyclic variations within relatively short time periods can be controlled by fixing constraints on the schedule of collection if external causes of variability are known. Alternatively, repeated samples from the same subject can be obtained in a short time period and pooled in an 'average' sample. To assess the optimal number of repeats per subject, within the time unit it is necessary to have some quantitative knowledge of both intra-individual and interindividual variance (Armstrong *et al.*, 1992). Transitional studies to describe these aspects of the marker in the population studied are therefore necessary when planning a prospective investigation. On the other hand, systematic trends over time are likely to be relevant to the carcinogenic process, and therefore variations over long time periods should be studied as independent factors and not as sources of random misclassification error.

In studies where external exposures have been determined at the individual level and correlated with specific adducts in the same individuals, e.g. number of cigarettes smoked versus aromatic DNA adducts in lung tissue (Phillips *et al.*, 1988), there is still a high degree of interindividual variation in the amount of adduct formed for a given exposure. This interindividual variation can result, for example, from differences in carcinogen metabolism, and some studies have identified an impact of these effect modifiers on the dose–response relationship between exposure and adducts. For example, Vineis *et al.* (1994) reported that, among cigarette smokers, those with a slow acetylator phenotype have higher average 4-ABP–Hb adduct levels than rapid acetylators and that this effect was particularly evident at low exposure levels. Adduct assessment can therefore find a major application in applied (transitional) studies trying to elucidate proposed mechanisms of action for a defined carcinogen. Lang *et al.* (1994), for example, studied the interaction between P4501A2 and NAT-2 polymorphism and a preference for well-cooked meat, as assessed by questionnaire in a case–control study of colon cancer. A pathway to link the intake of heterocyclic amines from cooked meat to the endogenous formation of carcinogenic *N*-acetoxyarylamine was proposed, and indeed the results corroborate the proposed metabolic pathway. High and low intakes of heterocyclic amines were indirectly estimated by a preference for well-cooked or lightly cooked meat. Increased excretion of heterocyclic amine has been shown in the urine of consumers of pan-fried meat, but DNA adduct levels, which should be the final products of the proposed mechanism, in consumers and non-consumers of well-done meat have not been described.

The above examples illustrate the use of adducts to test the impact of polymorphisms in the metabolism of carcinogens on one step of the carcinogenic pathway. As a consequence of such work, the rationale for the integration of genetic polymorphic markers for such genes into case-control studies can be strengthened.

The fact that adducts represent an integration of external exposure and interindividual variability in carcinogen metabolism, DNA repair, etc. makes them attractive because they may provide a more relevant measure of exposure, i.e. the biologically effective dose. However, it also means that the relationship between exposure and adduct level is not simple. As public health decisions are normally based on ambient exposure levels, more understanding of the relationship between these levels and adducts needs to be developed if the latter are to be used in making such decisions.

To complete this review, the possible applications of adducts in intervention studies to modify the risk of cancer or premalignant conditions are worth mentioning. In the context of the application of adducts, these can be divided into interventions that prevent exposure to genotoxic agents and those which tend to modify the metabolism of pro-carcinogens once exposure has occurred. In these cases, adducts would provide an early marker of the intervention efficacy and could be used in *ad interim* analyses for the purpose of monitoring of compliance. An example of this type of study is an ongoing short-term intervention to reduce aflatoxin-albumin adducts using the drug oltipraz (Bolton *et al.*, 1993). A different type of study is illustrated by Verhagen *et al.* (1995) who attempted to modulate the urinary levels of 8-oxodeoxyguanosine by increasing the dietary intake of fresh vegetables. This approach is an interesting one for testing hypotheses of mechanisms of carcinogenesis in humans. If adducts are to be used as surrogate end-points for disease outcome in intervention studies then considerably more understanding of the relationship between these events is required.

It should be also be noted that because of the ethical implications of any intervention, it is mandatory that the metabolic pathway leading to the production of adducts, together with its possible confounders and effect modifiers, is carefully described if adducts are to be used as outcome variables in intervention studies.

Acknowledgements

C.P.W. was partially supported by a grant from the NIEHS, USA, no. 1-P01 ES06052.

References

Aguilar, F., Hussain, S.P. & Cerutti, P. (1993) Aflatoxin B_1 induces the transversion of G→T in codon 249 of the p53 tumor suppressor gene in human hepatocytes. *Proc. Natl Acad. Sci. USA*, 90, 8586–8590

Armstrong, B.K., White, E. & Saracci, R., eds (1992) *Principles of Exposure Measurement in Epidemiology, Monographs in Epidemiology and Biostatistics*, Vol. 21, Oxford, Oxford Medical Publications

Autrup, H., Seremet, T., Wakhisi, J. & Wasunna, A. (1987) Aflatoxin exposure measured by urinary excretion of aflatoxin B1–guanine adduct and hepatitis B virus infection in areas with different liver cancer incidence in Kenya. *Cancer Res.*, 47, 3430–3433

Bär, W., Kratzer, A., Mächler, M. & Schmid, W. (1988) Postmortem stability of DNA. *Forensic Sci. Int.*, 39, 59–70

Bartsch, H. & Montesano, R. (1984) Relevance of nitrosamines to human cancer. *Carcinogenesis*, 5, 1381–1393

Belinsky, S.A., White, C.M., Devereux, T.R., Swenberg, J.A. & Anderson, M.W. (1987) Cell selective alkylation of DNA in rat lung following low dose exposure to the tobacco specific carcinogen 4-(*N*-methyl-*N*-nitrosamino) -1-(3-pyridyl)-1-butanone. *Cancer Res.*, 47, 1143–1148

Belinsky, S.A., Foley, J.F., White, C.M., Anderson, M.W. & Maronpot, R.R. (1990) Dose–response relationship between O^6-methylguanine formation in Clara cells and induction of pulmonary neoplasia in the rat by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Cancer Res.*, 50, 3772–3780

Bessho, T., Roy, R., Yamamoto, K., Kasai, H., Nishimura, S., Tano, K. & Mitra, S. (1993) Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase. *Proc. Natl Acad. Sci. USA*, 90, 8901–8904

Bianchini, F. & Wild, C.P. (1994a) 7-Methyldeoxyguanosine as a marker of exposure to environmental methylating agents. *Toxicol. Lett.*, 72, 175–184

Bianchini, F. & Wild, C.P. (1994b) Comparison of 7-medG formation in white blood cells, liver and target organs in rats treated with methylating carcinogens. *Carcinogenesis*, 15, 1137–1141

Bianchini, F., Montesano, R., Shuker, D.E., Cuzick, J. & Wild, C.P. (1993) Quantification of 7-methyldeoxyguanosine using immunoaffinity purification and HPLC with elec-

trochemical detection. *Carcinogenesis*, 14, 1677–1682

Bohr, V.A. (1991) Gene specific DNA repair. *Carcinogenesis*, 12, 1983–1992

Bolton, M.G., Munoz, A., Jacobson, L.P., Groopman, J.D., Maxuitenko, Y.Y., Roebuck, B.D. & Kensler, T.W. (1993) Transient intervention with oltipraz protects against aflatoxin-induced hepatic tumorigenesis. *Cancer Res.*, 53, 3499–3504

Campbell, T.C., Chen, J., Liu, C., Li, J. & Parpia, B. (1990) Non-association of aflatoxin with primary liver cancer in a cross-sectional ecologic survey in the People's Republic of China. *Cancer Res.*, 50, 6882–6893

Carmella, S.G. & Hecht, S.S. (1987) Formation of hemoglobin adducts upon treatment of F344 rats with the tobacco-specific nitrosamines 4- (methylnitrosamino)-1-(3-pyridyl)-1-butanone and N'-nitrosonornicotine. *Cancer Res.*, 47, 2626–2630

Chacko, M. & Gupta, R.C. (1988) Evaluation of DNA damage in the oral mucosa of tobacco users and non-users by 32P-adduct assay. *Carcinogenesis*, 9, 2309–2313

Chakravarti, D., Pelling, J.C., Cavalieri, E.L. & Rogan, E.G. (1995) Relating aromatic hydrocarbon-induced DNA adducts and c-H-*ras* mutations in mouse skin papillomas: The role of apurinic sites. *Proc. Natl Acad. Sci. USA*, 92, 10422–10426

Claycamp, H.G. (1992) Phenol sensitization of DNA to subsequent oxidative damage in 8-hydroxyguanine assays. *Carcinogenesis*, 13, 1289–1292

Coggon, D., Pannett, B., Acheson, E.D. (1984) Use of job-exposure matrix in an occupational analysis of bladder cancers on the basis of deaths certificates. *J. Natl Cancer Inst.*, 72, 61–65

Cohen, S.M. & Ellwein, L.B. (1990) Proliferative and genotoxic cellular effects in 2-acetylaminofluorene bladder and liver carcinogenesis: biological modeling of the ED01 study. *Toxicol. Appl. Pharmacol.*, 104, 79–93

Douki, T., Delatour, T., Bianchini, F. & Cadet, J. (1996) Observation and prevention of an artefactual formation of oxidized DNA bases and nucleosides in the GC-EIMS method. *Carcinogenesis*, 17, 347–353

Essigmann, J.M., Croy, R.G., Nadzan, A.M., Busby, W.F., Jr, Reinhold, V.N., Buchi, G. & Wogan, G.N. (1977) Structural identification of the major DNA adduct formed by aflatoxin B1 in vitro. *Proc. Natl Acad. Sci. USA*, 74, 1870–1874

Fischer, W.H. & Lutz, W.K. (1995) Correlation of individual papilloma latency time with DNA adducts, 8-hydroxy-2'-deoxyguanosine, and the rate of DNA synthesis in the epidermis of mice treated with, 7, 12-dimethylbenzo[*a*]anthracene. *Proc. Natl Acad. Sci. USA*, 92, 5900–5904

Goldberg, M., Kromhout, H., Guenel, P., Fletcher, A.C., Gerin, M., Glass, D.C., Heederik, D., Kauppinen, T. & Ponti, A. (1993) Job exposure matrices in industry. *Int. J. Epidemiol.*, 22, S10–15

Greenblatt, M.S., Bennett, W.P., Hollstein, M. & Harris, C.C. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.*, 54, 4855–4878

Groopman, J.D. (1994) Molecular dosimetry methods for assessing human aflatoxin exposures. In: Eaton, D.A. & Groopman, J.D, eds, *The Toxicology of Aflatoxins, Human Health, Veterinary and Agricultural Significance*. San Diego, Academic Press, pp. 259–279

Groopman, J.D., Hasler, J.A., Trudel, L.J., Pikul, A., Donahue, P.R. & Wogan, G.N. (1992) Molecular dosimetry in rat urine of aflatoxin-*N*7-guanine and other aflatoxin metabolites by multiple monoclonal antibody affinity chromatography and immunoaffinity/high performance liquid chromatography. *Cancer Res.*, 52, 267–274

Groopman, J.D., Wild, C.P., Hasler, J., Junshi, C., Wogan, G.N. & Kensler, T.W. (1993) Molecular epidemiology of aflatoxin exposures: validation of aflatoxin-*N*7-guanine levels in urine as a biomarker in experimental rat models and humans. *Environ. Health Perspect.*, 99, 107–113

Hall, A.J. & Wild, C.P. (1992) Aflatoxin biomarkers [letter; comment]. *Lancet*, 339, 1413–1414

Hall, A.J. & Wild, C.P. (1994) Epidemiology of aflatoxin-related disease. In: Eaton, D.A. & Groopman, J.D., eds, *The Toxicology of Aflatoxins, Human Health, Veterinary and Agricultural Significance*. San Diego, Academic Press, pp. 233–258

Hammond, S.K., Coghlin, J., Gann, P.H., Paul, M., Taghizadeh, K., Skipper, P.L. & Tannenbaum, S.R. (1993) Relationship between environmental tobacco smoke exposure and carcinogen-hemoglobin adduct levels in nonsmokers. *J. Natl Cancer Inst.*, 85, 474–478

Hemminki, K., Dipple, A., Shuker, D.E.G., Kadulbar, F.F., Segerbäck, D. & Bartsch, H., eds (1994) *DNA Adducts, Identification and Biological Significance* (IARC Scientific Publications No. 125). Lyon, International Agency for Research on Cancer

Hertzog, P.J., Smith, J.R. & Garner, R.C. (1982) Characterisation of the imidazole ring-opened forms of trans-8,9-dihydro-8,9-dihydro-8-(7-guanyl)9-hydroxy aflatoxin B1. *Carcinogenesis*, 3, 723–725

Hussain, S.P., Aguilar, F., Amstad, P. & Cerutti, P. (1994) Oxy-radical induced mutagenesis of hotspot codons 248

and 249 of the human p53 gene. *Oncogene*, 9, 2277–2281

Izzotti, A., Bagnasco, M., Scatolini, L., Rovida, A. & De Flora, S. (1993) Post-mortem stability of benzo[*a*]pyrene diolepoxide–DNA adducts in rat organs. *Carcinogenesis*, 14, 2185–2187

Kat, A.G. & Thilly, W.G. (1994) Mutational spectra of endogenous genes in mammalian cells. In: Hemminki, K., Shuker, D.E.G., Kadlubar, F.F., Segerbäck, D. & Bartsch, H., eds, *DNA Adducts, Identification and Biological Significance* (IARC Scientific Publications No. 125). Lyon, International Agency for Research on Cancer. pp. 371–383

Lang, N.P., Butler, M.A., Massengill, J., Lawson, M., Stotts, R.C., Hauer Jensen, M. & Kadlubar, F.F. (1994) Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiol. Biomarkers. Prev.*, 3, 675–682

Maclure, M., Bryant, M.S. & Skipper, P.L.A. (1990) Decline of the hemoglobin adduct of 4-aminobiphenyl during withdrawal from smoking. *Cancer Res.*, 50, 181–184

Maher, V.M., Yang, J-L., Mah, C-M. & McCormick, J.J. (1989) Comparing the frequency and spectra of mutations induced when and SV-40 based shuttle vector containing covalently bound residues of structurally-related carcinogens replicates in human cells. *Mutat. Res.*, 220, 83-92

Male, R., Haukanes, B.I., Helland, D.E. & Kleppe, K. (1987) Substrate specificity of 3-methyladenine-DNA glycosylase from calf thymus. *Eur. J. Biochem.*, 165, 13–19

Margison, G.P. & O'Connor, P.J. (1979) Nucleic acid modifications by N-nitroso compounds. In: Grover, P.L., ed., *Chemical Carcinogens and DNA*, Vol. 1. Boca Raton, FL, CRC Press, pp. 111–159

Martin, C.N. & Garner, R.C. (1977) Aflatoxin B-oxide generated by chemical or enzymic oxidation of aflatoxin B_1 causes guanine substitution in nucleic acids. *Nature*, 267, 863–865

May, A., Nairn, R.S., Okumoto, D.S., Wassermann, K., Stevnsner, T., Jones, J.C. & Bohr, V.A. (1993) Repair of individual DNA strands in the hamster dihydrofolate reductase gene after treatment with ultraviolet light, alkylating agents, and cisplatin. *J. Biol. Chem.*, 268, 1650–1657

Mooney, L.A., Santella, R.M., Covey, L., Jeffrey, A.M., Bigbee, W., Randall, M.C., Cooper, T.B., Ottman, R., Tsai, W-Y., Wazneh, L., Glassman, AH., Young, T-L. & Perera, F.P. (1995) Decline of DNA damage and other biomarkers in peripheral blood following smoking cessation. *Cancer Epidemiol. Biomarkers Prev.*, 4, 627–634

Mustonen, R. & Hemminki, K. (1992) 7-Methylguanine levels in DNA of smokers' and non-smokers' total white blood cells, granulocytes and lymphocytes. *Carcinogenesis*, 13, 1951–1955

Ochiai, M., Watanabe, M., Kushida, H., Wakabayashi, K., Sugimura, T. & Nagao, M. (1996) DNA adduct formation, cell proliferation and aberrant crypt focus formation induced by PhIP in male and female rat colon with relevance to carcinogenesis. *Carcinogenesis*, 17, 95–98

Ozbal, C.C., Velic, I., Soohoo, C.K., Skipper, P.L. & Tannenbaum, S.R. (1994) Conservation of histone carcinogen adducts during replication: implications for long-term molecular dosimetry. *Cancer Res.*, 54, 5599–5601

Ozturk, M. (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*, 338, 1356–1359

Pegg, A.E. (1984) Methylation of the O6 position of guanine in DNA is the most likely initiating event in carcinogenesis by methylating agents. *Cancer Invest.*, 2, 223–231

Perera, F., Jeffrey, A., Santella, R.M., Brenner, D., Mayer, J., Latriano, L., Smith, S., Young, T.L., Tsai, W.Y., Hemminki, K. & Brandt-Rauf, P. (1990) Macromolecular adducts and related biomarkers in biomonitoring and epidemiology of complex exposures. In: Vainio, H., Sorsa, M. & McMichael, A.J., eds, *Complex Mixtures and Cancer Risk* (IARC Scientific Publication No. 104). Lyon, International Agency for Research on Cancer, pp. 164–180

Perera, F.P., Hemminki, K., Gryzbowska, E., Motykiewicz, G., Michalska, J., Santella, R.M., Young, T.L., Dickey, C., Brandt-Rauf, P. & DeVivo, I. (1992) Molecular and genetic damage in humans from environmental pollution in Poland. *Nature*, 360, 256–258

Phillips, D.H., Hewer, A., Martin, C.N., Garner, R.C. & King, M.M. (1988) Correlation of DNA adduct levels in human lung with cigarette smoking. *Nature*, 336, 790–792

Phillips, D.H., Schoket, B., Hewer, A., Bailey, E., Kostic, S. & Vincze, I. (1990) Influence of cigarette smoking on the levels of DNA adducts in human bronchial epithelium and white blood cells. *Int. J. Cancer*, 46, 569–575

Poirier, M.C. & Beland, F.A. (1992) DNA adduct measurements and tumor incidence during chronic carcinogen exposure in animal models: implications for DNA adduct-based human cancer risk assessment. *Chem. Res. Toxicol.*, 5, 749–755

Poirier, M.C., Fullerton, N.F., Smith, B.A. & Beland, F.A. (1995) DNA adduct formation and tumorigenesis in mice during the chronic administration of 4-aminobiphenyl

at multiple dose levels. *Carcinogenesis*, 16, 2917–2921

Qian, G.S., Ross, R.K., Yu, M.C., Yuan, J.M., Gao, Y.T., Henderson, B.E., Wogan, G.N. & Groopman, J.D. (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China [see comments]. *Cancer Epidemiol. Biomarkers. Prev.*, 3, 3–10

Richardson, F.C., Beauchamp, R.O., Jr & Swenberg, J.A. (1987) Properties and biological consequences of alkylpyrimidine deoxyribonucleosides. *Pharmacol. Ther.*, 34, 181–213

Ross, R.K., Yuan, J.M., Yu, M.C., Wogan, G.N., Qian, G.S., Tu, J.T., Groopman, J.D., Gao, Y.T. & Henderson, B.E. (1992) Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet*, 339, 943–946

Rothman, N., Stewart, W.F. & Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers. Prev.*, 4, 301–311

Scates, D.K., Spigelman, A.D. & Venitt, S. (1995) Appearance of artefacts when using ^{32}P-postlabelling to investigate DNA adduct formation by bile acids *in vitro*: lack of evidence for covalent binding. *Carcinogenesis*, 16, 1489–1491

Shen, H.M., Shi, C.Y., Lee, H.P. & Ong, C.N. (1994) Aflatoxin B1-induced lipid peroxidation in rat liver. *Toxicol. Appl. Pharmacol.*, 127, 145–150

Shen, H.M., Ong, C.N., Lee, B.L. & Shi, C.Y. (1995) Aflatoxin B1-induced 8-hydroxydeoxyguanosine formation in rat hepatic DNA. *Carcinogenesis*, 16, 419–422

Shuker, D.E.G. & Farmer, P.B. (1992) Relevance of urinary DNA adducts as markers of carcinogen exposure. *Chem. Res. Toxicol.*, 5, 450–460

Skipper, P.L. & Tannenbaum, S.R. (1990) Protein adducts in the molecular dosimetry of chemical carcinogens. *Carcinogenesis*, 11, 507–518

Skipper, P.L., Peng, X., Soohoo, C.K. & Tannenbaum, S.R. (1994) Protein adducts as biomarkers of human carcinogen exposure. *Drug Metab. Rev.*, 26, 111–124

Srivatanakul, P., Parkin, D.M., Jiang, Y.Z., Khlat, M., Kao-Ian, U.T., Sontipong, S. & Wild, C.P. (1991) The role of infection of *Opisthorchis viverrini*, hepatitis B virus and aflatoxin exposure in the aetiology of liver cancer in Thailand. *Cancer*, 68, 2411–2417

Stone, J.G., Jones, N.J., McGregor, A.D. & Waters, R. (1995) Development of a human biomonitoring assay using buccal mucosa: comparison of smoking-related DNA adducts in mucosa *versus* biopsies. *Cancer Res.*, 55, 1267–1270

Szyfter, K., Hemminki, K., Szyfter, W., Szmeja, Z., Banaszewski, J. & Yang, K. (1994) Aromatic DNA adducts in larynx biopsies and leukocytes. *Carcinogenesis*, 15, 2195–2199

Tang, D., Santella, R.M., Blackwood, A.M., Young, T.-L., Mayer, J., Jaretzki, A., Grantham, S., Tsai, W.-Y. & Perera, F.P. (1995) A case–control molecular epidemiological study of lung cancer. *Cancer Epidemiol. Biomarkers Prev.*, 4, 341–346

Tornqvist, M. (1990) Formation of reactive species that lead to hemoglobin adducts during storage of blood samples. *Carcinogenesis*, 11, 51–54

Trottier, Y., Waithe, W.I. & Anderson, A. (1992) Kinds of mutations induced by aflatoxin B1 in a shuttle vector replicating in human cells transiently expressing cytochrome P4501A2 cDNA. *Mol. Carcinog.*, 6, 140–147

van Schooten, F.J., Hillebrand, M.J., van Leeuwen, F.E., van Zandwijk, N., Jansen, H.M., Den Engelse, L. & Kriek, E. (1992) Polycyclic aromatic hydrocarbon-DNA adducts in white blood cells from lung cancer patients: no correlation with adduct levels in lung. *Carcinogenesis*, 13, 987–993

Verhagen, H., Poulsen, H.E., Loft, S., van-Poppel, G., Willems, M.I. & van Bladeren, P.J. (1995) Reduction of oxidative DNA-damage in humans by brussels sprouts. *Carcinogenesis*, 16, 969–970

Vineis, P., Bartsch, H., Caporaso, N., Harrington, A.M., Kadlubar, F.F., Landi, M.T., Malaveille, C., Shields, P.G., Skipper, P., Talaska, G. & Tannenbaum, S.R. (1994) Genetically based N-acetyltransferase metabolic polymorphism and low-level environmental exposure to carcinogens. *Nature*, 369, 154–156

Widlak, P., Bykov, V.J. & Hemminki, K. (1995) Formation of UV-photoadducts durings DNA purification. *Mutat. Res.*, 347, 117–119

Wiencke, J.K., Kelsey, K.T., Varkonyi, A., Semey, K., Wain, J.C., Mark, E. & Christiani, D.C. (1995) Correlation of DNA adducts in blood mononuclear cells with tobacco carcinogen-induced damage in human lung. *Cancer Res.*, 55, 4910–4914

Wild, C.P., Fortuin, M., Donato, F., Whittle, H.C., Hall, A.J., Wolf, C.R. & Montesano, R. (1993) Aflatoxin, liver enzymes, and hepatitis B virus infection in Gambian children. *Cancer Epidemiol. Biomarkers. Prev.*, 2, 555–561

Wogan, G.N. (1988) Detection of DNA damage in studies on cancer etiology and prevention. In: Bartsch, H., Hemminki, K. & O'Neill, I.K., eds, *Methods for Detecting DNA Damaging Agents in Humans: Applications in Cancer Epidemiology and Prevention* (IARC Scientific Publications No. 89). Lyon, International Agency for Research on Cancer, pp. 32–51

Wogan, G.N. (1992) Molecular epidemiology in cancer risk assessment and prevention: recent progress and avenues for future research. *Environ. Health Perspect.*, 98, 167–178

Zhang, Y.J., Chen, C.J., Lee, C.S., Haghighi, B., Yang, G.Y., Wang, L.W., Feitelson, M. & Santella, R. (1991) Aflatoxin B1-DNA adducts and hepatitis B virus antigens in hepatocellular carcinoma and non-tumorous liver tissue. *Carcinogenesis*, 12, 2247–2252

Corresponding author

C.P. Wild

Molecular Epidemiology Unit, Research School of Medicine, Algernon Firth Building, University of Leeds, Leeds LS2 9JT, UK

IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Somatic cell mutations in cancer epidemiology

R.J. Albertini and R.B. Hayes

Somatic cell mutations arising *in vivo* in reporter genes and in cancer-associated genes may now be measured in humans. Background mutation levels and mutational responses following various mutagen exposures are reviewed in this chapter. The detection methods are compared for similarities and differences based on the underlying biology of the systems. Currently available data on molecular mutational spectra are reviewed and the utility of such information is discussed in terms of mutagen exposure characterization and for defining the mutagenic basis of carcinogenesis. In addition to the reporter gene assays, recently developed assays for mutation in cancer-associated genes are considered. The strengths and limitations of using somatic cell mutations for cancer epidemiology and areas for future research are discussed.

Somatic cell mutations occur regularly and universally in all humans. Some are 'naturally occurring', arising continuously as spontaneous replication errors or in response to endogenous mutagens or DNA metabolism, while others are induced by external mutagens which are ubiquitous in the environment. Mutations are central to human carcinogenesis, and, consequently, exposures that cause mutations are also suspect of causing cancer. In principle, subjects who have experienced greater exposure to genotoxic agents will accumulate greater numbers of mutations, so that quantitation of the frequency of mutational events serves to identify heavily exposed individuals. In addition, specific genotoxic agents may produce specific types of mutations, implying that characterization of mutational errors (mutational spectrum) will be useful for identifying exposure to specific agents.

Modern in-vivo somatic mutation studies may employ any of seven assays for measuring in-vivo mutations in five different reporter genes, i.e. not functionally related to carcinogenesis. There are also methods for detecting mutations in cancer genes (oncogenes and tumour suppressor genes). Extensive databases are available for two of the reporter gene assays (*GPA* and *hprt*), while limited information is available for all others. Here we describe the genetic bases and methods for the major assays, summarize the results of applications in humans, and discuss the uses and limitations of the methods for future epidemiological studies.

Reporter gene somatic cell mutation assays: genetic basis and methods

The haemoglobin (Hb) genes

Genetic basis (Table 1). Hb is a tetrameric protein that constitutes >99% of the protein in non-nucleated mature red blood cells (RBCs) (reviewed in Stamatoyannapoulos et al., 1984). Several polypeptide chains, encoded by genes at several linked loci, constitute the intact Hb molecule, depending on the stage of development of the individual. Following birth, adult HbA ($\alpha_2\beta_2$) constitutes >95% of the Hb in mature RBCs, with a small amount of HbA2 ($\alpha_2\delta_2$) also present. The ε,γ (two loci) δ and β Hb genes are on chromosome 11p; the ζ and α (two loci) Hb genes are on chromosome 16p (Deisseroth et al., 1977, 1978).

Three classes of mutations affect the Hb genes. Two, the thalassaemias and hereditary persistence of fetal Hb (HPFH), produce abnormal levels of Hb polypeptides. These are not useful for mutational studies however, as non-genetic influences can reduce Hb levels in RBCs (Stamatoyannapoulos *et al.*, 1984). The third class of Hb mutations produces structural alterations in Hb. Over 500 such changes have been described, most being single amino acid changes in the α or the β chain. Of these, the best known is an A to T transversion in the gene that gives 'sickle cell' haemoglobin, or HbS. Although the gene is autosomal and contains three exons spanning 2 kb, the effective mutational target size for single base changes is a

Table 1. Somatic reporter gene mutations in humans: the reporter genes

Cells	Gene(s)	Chromosome	Gene size	Target size
RBC	Hb gene	Autosome, 11p	3 exons, 2 kb	bp (small)
	Hb gene	Autosome, 16p	3 exons, 2 kb	bp (small)
RBC	*GPA*	Autosome, 4q	7 exons, 44 kb	>44 kb (large)
T-lymphocytes	*HPRT*	X-linked, Xq	9 exons, 44 kb	>44 kb (large)
T-lymphocytes	*HLA-A*	Autosome, 6p	7 exons, 5 kb	>5 kb (large)
T-lymphocytes	TCR	Autosome, 14q	Large 'multi-gene'	Probably large
	TCR	Autosome, 7q	Large 'multi-gene'	Probably large

single base pair. The assay is thus limited to the detection of point mutations, which limits its practical sensitivity and restricts to a narrow range the spectrum of mutational mechanisms detected.

Hb variant frequency (VF) assays (Table 2). G.P. Stamatoyannopoulos and co-workers originally developed an assay based on highly specific antibodies that discriminate the different mutant haemoglobins from each other and from normal haemoglobin, requiring samples of 1 ml blood (reviewed in Stamatoyannopoulos et al., 1984). Scoring techniques for rare variant cells with a mutated Hb are based on fluorescence labelling of fixed RBCs on slides. As haemoglobin is an intracellular protein, fixation and permeabilization are required for antibody entrance. Recent refinements have included newer methods for permeabilizing and labelling RBCs in suspension for automated cell sorting and counting (Bigbee et al., 1981). More recently, automated microscopy has been used to screen large numbers of cells on slides (Tates et al., 1989). One or more structural Hb mutations may be scored in an assay.

Although there have been no systematic descriptions of all technical sources of error in this assay, quality of RBCs and extent of permeabilization could be major sources, as could the quality of the immunological reagents and non-specific staining. Poisson counting of rare events produces relatively large errors.

The GPA gene

Genetic basis (Table 1) Glycophorin-A (GPA) is a polymorphic glycoprotein on RBC surfaces present at approximately 5×10^5 molecules/cell (Furthmayer, 1977; Gahmberg *et al.*, 1979). The *GPA* gene on chromosome 4q spans 44 kb and contains seven exons (Kudo & Fukuda, 1989). In principle, it is a large target for mutation. It has two codominantly expressed (M and N) alleles (Furthmayer, 1978) (except for rare nulls) which have approximately equal frequencies in all populations (Cartron *et al.*, 1990). Thus, approximately 50% of humans are M/N heterozygotes with both forms of the glycoprotein on RBC surfaces. The M and the N molecules differ by two non-adjacent amino acids and cannot be interconverted by simple point mutations.

GPA VF assays (Table 2). The currently used *GPA* assay measures the frequency of variant cells that have lost expression of the M form in blood samples from heterozygous individuals (M/N) (Langlois *et al.*, 1990; Jensen & Bigbee, 1996). A small (<1 ml) sample of blood is treated soon after blood draw to fix spherical erythrocytes which are then kept at 4°C until analysis. Variant cells are detected by flow cytometry with distinguishable fluorescent-labelled monoclonal antibodies specific for the N and M forms and counting variant cells that bind the anti-N but not the anti-M antibody. Simple loss mutations, many representing 'point mutations', are expressed as rare 0/N cells (Grant & Bigbee, 1993). These are referred to as 'hemizygous variants'. More complex mutations are expressed as rare N/N cells on the M/N background, i.e. a loss of function of one allele with double expression of the other. These are thought to arise from mutational events that lead to cellular homozygosity (or loss or

Table 2. Somatic reporter gene mutations in humans: the assays

Gene and cell	Sample size	Method(s)	Time for assay	Potential for error
Hb in RBC	1 ml	Immunological staining	Month	RBC quality
		—Manual slides	Days	Permeabilization
		—Automated slides	Hours–days	Immunological reagents
		—Cytometry		Non-specific staining
				Poisson counting
GPA in RBC	1 ml	Immunological staining	Hours–days	RBC quality
		—Cytometry		Immunological reagents
				Non-specific staining
				Poisson counting
HPRT in T-cells	10–50 ml	Short-term		Cycling cells give phenocopies
		Autoradiography		Observer error
		Manual slides	Days–week	Cell culture factors
		Automated slides (potential)		
		BrdU incorporation with differential fluorescence		Non-specific staining changes
		Manual slides	Days–week	Poisson counting
		Automated slides	Days	Instrumentation for
		Cytometry (potential)	(Short)	automated assays
				Poisson counting
		Cloning	Weeks	Inverse correlations
		Selection in tissue culture		between CE and MF
				Observer error
				Reagents and culture conditions
				Poisson counting
HLA in T-cells	10–50 ml	Cloning	Weeks	Non-specific lack of
		Immunoselection (cytotoxicity)		immunocytotoxicity
		in vitro with cell culture		Observer error
		outgrowth		Reagents and culture conditions
				Immunological reagents and complement
				Poisson counting
	5–10 ml	Immunological staining		Non-specific failure to label
		Cytometry		immunological reagents
				Instrumentation
				Poisson counting
TCR in T-cells	Few ml	Immunological staining		Non-specific failure to label
		Cytometry		immunological reagents
				Instrumentation

heterozygosity), such as chromosome missegregation, somatic recombination or gene conversion (Grant & Bigbee, 1993). These variants are termed 'homozygous variants'. The rare 0/N and N/N cells that have lost expression of the M allele fall within prescribed areas of the cytogram. The variant frequency (VF) is defined as:

$$VF = \frac{\text{number of 0/N or N/N variant erythrocytes}}{\text{total number of red blood cells analysed}}$$

The quality and age of the RBCs in the blood sample can affect results, necessitating the rapid fixation of collected samples. Antibody lot, quality and preparation are potential sources of technical variability, as are changes in cytometer performance. Poisson counting of small numbers produces relatively large errors. A gold standard has recently been developed for the *GPA* assay (Jensen & Bigbee, 1996). Cryopreservation of formalin-fixed spherical RBCs in medium and DMSO, and storage at –80°C or –150°C maintain samples with stable VFs for up to 6 months. Reference standards are available for interlaboratory quality control and standardization.

The HPRT gene

Genetic basis (Table 1). The *HPRT* gene encodes the HPRT enzyme which is constitutively expressed but dispensable in virtually all mammalian cells. HPRT phosphoribosylates its normal substrates hypoxanthine and guanine for conversion to inosinic acid (Stout & Caskey, 1985) and is required to phosphoribosylate purine analogues such as 6-thioguanine (TG) to their cytotoxic forms (Albertini, 1985a ; Stout & Caskey, 1985). Cells with normal HPRT activity are susceptible to the cytotoxicity of TG and related agents; mutants are resistant, thus providing a basis for selection.

The *HPRT* gene is X-linked and expressed as a single copy in all cells. The gene spans 44 kb and contains nine exons (Patel *et al.*, 1984). A total of 55 kb of DNA including and surrounding this gene has been sequenced, making the region one of the best characterized for mutational studies. In principle, *HPRT* is a large target for mutation.

HPRT VF and mutant frequency (MF) assays (Table 2). There are two assays for assessing *in vivo HPRT* mutations in human T-lymphocytes. Both are based on the resistance of mutant cells to HPRT-dependent cytotoxicity of the purine analogue TG (Albertini *et al.*, 1990). Both assays require similar collection and fractionation protocols. Although the assays can be performed on 5–10 ml of blood, it is preferable to obtain 30 ml or more.

The VF assay (Table 2). The first assay to be developed was a short-term phenotypic assay which has the advantage of speed and potential for automation but the disadvantage of consuming the TG^r T-cells which are then not available for study (Strauss & Albertini, 1977, 1979; Stark *et al.*, 1984). As there is no way to unequivocally demonstrate the mutational basis of the TG^r cells, they are termed variants and their frequencies in blood are termed VFs. For assay, cryopreserved MNCs are thawed and stimulated with PHA in replicate short-term cultures, with or without TG, and incubated until culture termination at 24–30 h (cryopreservation is required to avoid labelling of normal T-cells that are in 'in cycle' *in vivo*; Albertini *et al.*, 1981). At termination, fixed cells are added in measured volumes to microscope slides, and then stained, autoradiographed and scored. Recent protocol modifications use BrdU staining and scoring by differential fluorescence (Ostrosky-Wegman *et al.*, 1988), approaches suitable for automation using either cell cytometry or image analysers.

There are several technical sources of variability in the short-term autoradiographic assay. Only rare labelled cells are counted, so Poisson errors inherent in counting small numbers can be large. The slide-based method is laborious and susceptible to observer error. Automated assays will have concerns with instrumentation. T-cells that are cycling *in vivo* may become labelled and be scored as mutants, unless measures such as cryopreservation are employed to remove this effect (Albertini *et al.*, 1981).

The MF assay (Table 2). The second *HPRT* assay depends on direct *in vitro* cloning in TG and, although laborious, allows for mutant isolation, in-vitro propagation and molecular analyses (Albertini *et al.*, 1982; Morley *et al.*, 1985; Henderson *et al.*, 1986; O'Neill, *et al.*, 1987). Fresh or thawed, cryopreserved MNCs are washed and plated in culture medium for direct cloning in the presence or absence of TG. (Cryopreservation is not required here to remove 'phenocopies' as it is in the autoradiographic assay, but it is convenient.)

A mutant frequency (MF) (and its confidence interval) is calculated from the ratio of the cloning efficiency of T-cells in the presence of TG to the cloning efficiency in its absence. There are large methodological differences between the two kinds of *HPRT* assays, which may give somewhat different results.

There are several sources of technical variability in the cloning assay (Robinson *et al.*, 1993). All investigators have reported a strong inverse correlation between control or non-selected cloning efficiencies and calculated mutant frequencies; observer error with failure to recognize slow-growing mutant colonies may result in variability. Attention to reagents and conditions is critical. Fortunately, the ability to cryopreserve MNCs for subsequent testing provides a protocol standard to control for intralaboratory drift.

The HLA gene

Genetic basis (Table 1). The several linked *HLA* loci include two classes of genes that encode cell surface recognition or restriction molecules of importance for antigen presentation in immune responses (Bodmer, 1984; Janeway & Travers, 1994). These genes constitute the major histocompatibility complex (MHC) in humans. Although the loci are extremely polymorphic, some alleles are present in a high proportion of individuals, i.e. approximately 50% of the population is heterozygous for either the *HLA-A2* or the *HLA-A3* allele.

The *HLA* complex is on chromosome 6p. In practice, mutation studies have been confined to the *HLA-A* gene for primary detection, although molecular studies have defined loss of other linked genes. The *HLA-A* gene is autosomal, spans 5 kb and contains seven exons; it is a large target for mutation.

HLA mutant frequency and variant frequency assays (Table 2). There are also two assays for assessing *HLA* mutations in human T-cells. The constitutional *HLA* genetic background of an individual must be known to measure somatic mutations at this locus, i.e. an individual must normally express one of the test HLA antigens in heterozygous form. The assays measure cellular loss of one codominantly expressed antigen.

Cloning (MF) assay (Table 2). In the cloning assay, (Janatipour *et al.*, 1988; McCarron *et al.*, 1989), peripheral blood is obtained and the MNC fraction is separated. As for *HPRT*, this assay involves inoculation of MNCs into the wells of microtitre plates in limiting dilutions. Selection is due to cytotoxicity by a relevant antibody in the presence of complement. Following selection, cells are directly inoculated into microtitre plates essentially as in the *HPRT* cloning assay. Colonies are scored at 16–20 days by inverted phase microscopy.

One major source of technical variability in the *HLA* assay is lack of killing by specific antibody. However, these 'phenocopies' can be recognized by testing growing colonies for resistance to the specific selecting antibody. The other sources of technical variability are inherent in cell culturing. The sources of immunological reagents and complement and the conditions of immunoselection can also produce technical variation. It should be possible to produce a cryopreserved cell standard as described for *HPRT*.

VF assay (Table 2). A short-term flow cytometry assay has also been described as measuring *HLA* loss mutations in T-cells (Kushiro *et al.*, 1992). The method uses specific biotin-labelled anti-HLA antibodies to label *HLA* gene products on T-cells, FITC-conjugated monoclonal anti-CD3 antibody to label all T-cells, and two-colour cytometric analysis to score CD3+ T-cells lacking the target HLA antigens.

The rare cells that have lost the target HLA antigen, i.e. less than 1/25 × that of normal HLA+ cells, fall within a prescribed area of the cytogram. The variant frequency is defined as:

$$VF = \frac{\text{number of CD3+ target HLA antigen-lacking lymphocytes}}{\text{total of CD3+ lymphocytes}}$$

Sources of error in the assay are non-specific failure to label cells, immunological reagents, instrumentation and Poisson sampling. When used only in the cytometry mode, the assay may be subject to serious phenocopy errors. However, cell sorting can and has been employed to clone and propagate single variant cells *in vitro* for molecular analyses. In principle, the genetic basis of all variants detected in this assay could be verified by molecular analyses.

The TCR gene

Genetic basis (Table 1). There are at least four T-cell receptor (TCR) genes, i.e. the α, the β, the γ and the

δ genes, located on chromosomes 14q (α and, within it, δ), 7q (β) and 7p (γ). Germ-line TCR genetic segments consist of variable (V), joining (J), diversity (D) (β and γ TCR genes only) and constant (C) regions (Janeway & Travers, 1994). Rearrangements occur during the differentiation of T-cells and are mediated by a recombinase system, termed V(D)J recombinase, through which the different V, D and J regions of any TCR germ-line gene are joined in all possible combinations. This confers the TCR gene uniqueness found in mature T-cells. The rearranged genes characterize and identify a specifically reactive T-cell and its clonal descendants.

The complete TCR on the T-cell surface consists of a constant molecule, termed CD3, and the TCR heterodimeric molecule consisting of either an α and β or a γ and δ polypeptide chain, encoded by the respective TCR genes (Clevers *et al.*, 1988; Kyoizumi *et al.*, 1990). A given T-cell expresses either the α/β heterodimeric TCR or the γ/δ heterodimeric TCR—never both. More than 90% of the peripheral blood T-cells express the α/β TCR. The TCR gene mutational assay currently in use focuses only on CD4+/TCRα/βT-cells.

If one of the molecules of the TCR heterodimer is defective for any reason, the CD3TCRα/βTCR complex fails to form on the T-cell surface, and the CD3 molecule accumulates in the cytoplasm. Allelic exclusion operates for the TCR genes, rendering any cell functionally hemizygous for gene expression. Therefore, mutational loss of gene function is not masked by a second allele.

TCR variant T-cells are recognized by the absence of the CD3 molecule from the surface of a CD4+ T-cell. The presumed mutation arises in one of the TCR genes (α or β). (As allelic exclusion is not operative for the CD3 gene, mutational loss of function would require somatic mutation of both alleles at this locus.)

The TCR genes are large genetic segments when in the germline configuration. However, somatic mutations may arise in rearranged genes, which are much smaller. It is not possible at this time to define a target size for the TCR gene mutations.

The TCR VF assay (Table 2). The TCR gene mutation assay is also based on analysis by flow cytometer (Kyoizumi *et al.*, 1992). Commercially available anti-CD3 is phycoerythrin-labelled and anti-CD4 is FITC-labelled, allowing for double labelling of normal T-cells. Variant cells are CD4+ cells that have lost expression of CD3. VFs are calculated as the number of CD3– cells, i.e. cells with CD3 expression level less than 1/25 × that of normal CD4 cells, divided by the total number of CD4+ T-cells.

Sources of technical variability for the TCR gene mutation assay are variabilities in immunological reagents and in flow cytometer performance. Furthermore, as detection of the cell surface molecules are the scored phenotype, the condition of the cells, surface perturbations, etc. could be potential sources of technical variability. Large numbers of variant cells are scored in this assay, and so errors inherent in counting small numbers should not be a problem. As it is not possible to clone most of the TCR gene variants *in vitro* for further analyses, the mutational basis of the measured phenotype is usually not demonstrable.

Reporter gene somatic mutations

General population studies

There is an extensive literature on in-vivo somatic mutations in humans, with extensive databases for *GPA* and *HPRT* mutations (reviewed in Albertini *et al.*, 1990; Cole & Skopek, 1994). This section presents representative results for in-vivo variant and mutant frequencies and molecular mutational spectra.

Hb variant frequency. The Hb mutation assay has been used in demonstration studies of very few subjects, showing in the original report a mean background VF of 11.0×10^{-8} (range $4–30 \times 10^{-8}$) for the single base changes producing the gene HbS and HbC mutations and, in a more recent report on five subjects, a lower mean background VF value of 3.7×10^{-8} (Stamatoyannopoulos & Nute, 1981; Bernini *et al.*, 1990). Intra-individual VF values have remained stable over relatively long time intervals, i.e. at least months, and smoking reportedly elevates the haemoglobin VF values approximately twofold.

GPA variant frequency. Hundreds to thousands of individuals have been studied for in-vivo *GPA* mutations using different versions of the assay (reviewed in Cole & Skopek, 1994). In general, mean hemizygous and homozygous VF values for adults have each averaged 10×10^{-6}, with the for-

mer somewhat lower and the latter somewhat higher. There is wide interindividual variability in VF. Values for newborns and children have been lower than in adults, indicating a clear age effect, with about a twofold increase by age 70. Smoking shows a weak association with increased VF, but systematic differences are not found by gender or ethnicity.

HPRT variant frequency. For the short-term tests, most *HPRT* T-cell VFs have been determined using the autoradiographic version of the assay. Mean values in normal populations range from 10^{-6} to 10^{-5} (reviewed in Albertini *et al.*, 1990; Cole & Skopek, 1994). VFs are increased in adults compared with newborns, and in adults they generally show an increase in frequency with age. Interindividual variability ranges from 10-fold to 30-fold and smoking has generally been associated with an increase in variants. Repeat sampling from individual donors has shown differences of twofold to fourfold in most instances, although differences as great as ninefold have been reported.

HPRT mutant frequency. The largest in-vivo somatic mutation database is for *HPRT* in T-cells, as determined by cloning assay (reviewed in Albertini *et al.*, 1990; Cole & Skopek, 1994). As these mutations are being confirmed by molecular studies, frequencies of TG^r cells are termed MF. A recent statistical analysis of four large data sets (Sussex, UK; Vermont, USA; Paris, France; and Leiden, the Netherlands), including 72 newborns, 70 children and 418 adults, showed ranges of MF values of $0.1–14.7 \times 10^{-6}$ for newborns, $0.5–39.5 \times 10^{-6}$ for children, and $0.8–81.7 \times 10^{-6}$ for adults, with a consistent trend from birth to old age (Robinson *et al.*, 1994). Smoking was clearly associated with increased MFs in the Sussex and Leiden data sets only, although a recent analysis in Vermont also showed a smoking effect. Neither sex nor ethnicity has been found to influence *HPRT* MFs.

HLA mutant frequency. As *HLA* mutations determined by cloning can be confirmed by molecular studies, frequencies of antigen loss cells are also termed mutant frequencies. Mean *HLA* MF values are $2–3 \times 10^{-5}$. *HLA* MFs increase with age, but the effects of tobacco use have not been investigated (reviewed in Albertini *et al.*, 1990; Kyoizumi *et al.*, 1992).

HLA variant frequency. *HLA* VFs determined by cytometric assay are at least five times higher than the clonally determined *HLA* MFs reported earlier, i.e. 1.5×10^{-4} and 0.7×10^{-4} for *HLA-A2* and *HLA-A24* loss variants, respectively (Kushiro *et al.*, 1992). (It is noteworthy that the anti-HLA-A2 antibody was the same for the cloning and cytometric assays.) The reasons for this and for the difference between VFs determined for the two *HLA-A* alleles are unknown.

TCR variant frequency. There are relatively few published reports of background TCR gene VFs. As this is a phenotypic assay, the CD3 loss T-cells are referred to as variants. The background VFs are remarkably high compared with other reporter genes (except *HLA* loss determined by cytometry), i.e. 2.5×10^{-4} (Kyoizumi *et al.*, 1990, 1992). The reasons for this are unknown, although speculation has centred on the role of this genetic region in immune diversification, which may require hypermutability (*HLA* also is involved in immune diversification, but only shows high frequencies of variants in the cytometric assay). Mean TCR VFs increase with age and are reported to be one- to fourfold higher in males. No published reports of studies on newborns or young children, or of smoking effects are available.

Studies in the inherited genetic instability

Marked increases of several VFs and MFs have been observed in patients who are homozygous for certain rare genetic instability syndromes due to defects in DNA repair (reviewed in Albertini *et al.*, 1990; Cole & Skopek, 1994). For example, ataxia telangiectasia (AT) homozygotes have shown increases in *GPA* hemizygous (0/N) and homozygous (N/N) VFs, *HPRT* MFs and TCR VFs, although no effect was seen for Hb VFs. Marked excesses in somatic cell mutations have generally also been found in Bloom's syndrome and Fanconi's anaemia (FA) patients. However, a recent study did not detect increases in *HPRT* MFs but did find *GPA* VF increases in FA (Sala-Trepat *et al.*, 1993). (An earlier study did show *HPRT* MF increases in FA; Vijayalaxmi *et al.*, 1985). For xeroderma pigmentosum (XP), increases were found for *HPRT* T-lymphocyte MFs, but not for *GPA* or Hb VFs. As the cause of genetic instability in XP is hypersensitivity to UV, this finding supports the occurrence of somatic mutations in T-cells as they circulate

throughout the body—in this case in skin, where they are subject to UV irradiation. By contrast, the RBC precursor cells, in which *GPA* and Hb mutations must occur, do not receive similar UV exposures. Therefore, somatic mutation appears linked to a single body compartment for the RBC precursors, but not for the T-cells.

Studies of environmental exposures to genotoxic agents (reviewed in Cole & Skopek, 1994)

Hb variant frequency. Increased HbS and HbC VFs were found among subjects exposed to X-rays, and Hb_{Leiden} VFs were increased after accidental exposure to ^{137}Cs (Stamatoyannopoulos & Nute, 1981; Bernini *et al.*, 1990). Another study found HbS VFs of 8, 18, 27 and 43×10^{-8} in four individuals following their exposure to ethylene oxide, compared with a background range in unexposed individuals of $0–8 \times 10^{-8}$ (Tates *et al.*, 1989). However, the three exposed individuals with the most elevated values were also smokers. There have been too few studies of Hb mutations following mutagen exposures to evaluate its performance.

GPA variant frequency (cytometric assay). By contrast to the Hb system, there is a wealth of data on *GPA* mutations following human mutagen exposures.

Radiation. Among the earliest population studies with the *GPA* system were those of atomic bomb survivors. Significant *GPA* mutation inductions with radiation exposures were reported in two studies, with the latter showing 63.0×10^{-6}/Gy, 32.0×10^{-6}/Gy and 0.14×10^{-6}/Gy for 0/N, 0/M and M/M variants, respectively (Langlois *et al.*, 1987; Kyoizumi *et al.*, 1989). (These were performed with an earlier version of the *GPA* assay that could also measure loss of the N form of *GPA*.) Similar but somewhat lower values for hemizygous variant increases have been reported more recently (Langlois *et al.*, 1993). All results are population means; in the absence of pre-exposure reference values, individual VFs could not be used to determine individual radiation exposures, as greatly elevated VFs were widely variable in high-dose subjects. These values were obtained more than 40 years after the mutagen exposure, indicating that the *GPA* marker can be long lived. This long memory and the wide interindividual variabilities in response are both consistent with the multipotent bone marrow stem cells being the mutational targets at the time of the bomb blasts. As this stem cell pool contains a limited number of cells, high-dose irradiation reaching it leaves few survivors and wide fluctuations of induced mutants among individuals. Although the numbers of variants among individuals will have an enormous range, the mean VFs in the population at the different radiation exposure levels will reflect the dose. Other studies, in Goiana, Brazil, and among victims of the Chernobyl accident, have reported similar results (Straume *et al.*, 1991; Jensen *et al.*, 1995).

By contrast to the atomic bomb results, patients receiving local irradiation to solid tumours (Hodgkin's disease, prostate) have shown no elevations of *GPA* VFs (Mendelsohn, 1990; Grant & Bigbee, 1994). The geometry of radiation exposures, as well as intensity and duration, appear to be important in determining the *GPA* mutational response. This result is consistent with the lack of VF elevations in XP patients noted above, in that both point to the bone marrow as the sole site of *in vivo* mutations. However, the results of a study in Japanese patients who had received the emitter Thorotrast (^{232}Th) in the 1930s and 1940s for radiographic visualization are somewhat at variance with this (Umeki *et al.*, 1991; Kyoizumi *et al.*, 1992). Although this agent accumulates in body tissues, there were no significant elevations of *GPA* VFs in 10 patients who, as a group, did show elevations of TCR gene T-cell VFs (see below). It is possible that the bone marrow received significantly lower doses than did other tissues.

Chemotherapy-related exposures. Earlier studies of a heterogeneous group of 30 cancer patients treated with a variety of cytotoxic agents and a more recent study of breast cancer patients receiving the combination CAF (cyclophosphamide, adriamycin, 5-fluorouracil) showed consistent elevations of 0/N VFs, while patients receiving CMF (methotrexate substituted for adriamycin) showed lower and more variable elevations (Bigbee *et al.*, 1990). The VF elevations in chemotherapy patients treated with these S-phase specific agents were transient with values returning to normal in several months, consistent with the half-life of RBCs (Bigbee *et al.*, 1990; Grant & Bigbee, 1994).

In children, chemotherapy post-treatment elevations were significant. By contrast with the adult studies, however, when plotted against time post-therapy, VFs for both the 0/N and N/N variants have remained significantly elevated for more than 10 years (Hewitt & Mott, 1992; Mott *et al.*, 1994). Radiation alone did not elevate *GPA* VF in the childhood studies.

Lastly, a study of platinum therapy in adult patients with germ cell tumours reported significant elevations of both 0/N and N/N VFs at different time points throughout therapy, with some significant increases persisting up to 6 months post-therapy (Perera *et al.*, 1992a; Grant & Bigbee, 1994) (*HPRT* T-cell mutations by cloning assay were reported as only marginally elevated in this study—see below).

Unlike localized radiotherapy, therefore, chemotherapy clearly induces *GPA* mutations, which most probably occur in the differentiated rather than the multipotent RBC precursors depending on the chemotherapeutic agent in question. The life spans of the resultant mutants are therefore some multiple of an RBC's life span, producing memories of months to years, depending on when in differentiation most mutations occur. In children there may be a tendency for mutations to occur in early progenitor cells, with some arising in the multipotent stem cells. The *GPA* memory for chemically induced mutations appears to be longer in children than in adults.

Other exposures. Measurements of *GPA* variants, *HPRT* mutant T-cells and DNA adducts were made in iron foundry workers exposed to polycyclic aromatic hydrocarbons (PAHs), and all results were related to ambient PAH exposure levels determined by personal and area monitoring (Perera *et al.*, 1993, 1994). 0/N hemizygous VFs increased slightly, but not significantly, with ambient exposures, while N/N VFs were unrelated to the exposures. Neither set of VFs correlated with PAH–DNA adducts (the *HPRT* mutations in this study are discussed below).

A study of 24 workers in China heavily exposed to benzene and 23 matched controls showed a significant increase in mean N/N VF (13.9×10^{-6} versus background of 7.4×10^{-6}) but no change in 0/N VFs in benzene-exposed workers, suggesting that benzene induces gene-duplication mutations in bone marrow stem cells, which may be the types of relevant pathogenic events in benzene-induced leukemias (Rothman *et al.*, 1995).

A recent study of reinforced plastics workers in Finland exposed to styrene revealed significantly elevated *GPA* N/N VFs among the most heavily exposed workers, particularly women (Bigbee *et al.*, 1996).

HPRT variant frequency (short term assays): radiation. Elevated *HPRT* VFs have been reported for individuals who received heavy accidental exposures to gamma irradiation from a ^{60}Co source in Mexico, for individuals in Kiev, Ukraine, at the time of the Chernobyl accident, and for subjects exposed to gamma irradiation from the ^{137}Cs source in Goiana, Brazil (Tates *et al.*, 1989; Ostrosky-Wegman *et al.*, 1990).

An early study reported radiation-induced elevations of *HPRT* VFs among 12 cancer patients who received 1.8–2.0 Gy/day localized radiotherapy for totals of 20–60 Gy, with the range of VF values in the treated individuals being higher than even the highest background values in control smokers. VFs returned towards normal, from 5 to 32 weeks post-treatment (Ammenheuser *et al.*, 1991).

Chemotherapy-related exposures. The earliest autoradiographic studies of *in vivo HPRT* mutations in humans were in cancer patients receiving cytotoxic chemotherapies (Albertini, 1985a, 1985b). However, the most detailed time-series study of induced *HPRT* VF elevations involved multiple sclerosis patients receiving i.v. bolus infusions of cyclophosphamide (750 mg/m^2 per month) (Ammenheuser *et al.*, 1988). Non-smoking and smoking patients were reported to have mean VFs before treatments of 1.52×10^{-6} and 6.56×10^{-6}, respectively. Two weeks after the first treatment, mean VFs were 29.07×10^{-6} ($n = 4$). All treated individuals showed elevated VFs, the lowest being 11.61×10^{-6}. At 4 weeks after the first treatment, the mean value fell to 5.58×10^{-6} ($n = 4$), and declined further to 3.50×10^{-6} ($n = 5$) at 7–13 weeks after the last treatment. Although clearly indicating a mutagenic effect, these rapid rises and falls in VFs suggest that cytotoxicity, cell division and its abatement may have had some effects at these high acute doses.

Other exposures. An early study of nurses who dispense chemotherapy showed no VF elevations over background (Albertini *et al.*, 1988). However, all safety practices had been observed. By contrast, another early study (using a version of the autoradiograpahic assay that did not take steps to eliminate phenocopies) did show a statistically significant (five- to sixfold) increase in VFs in cyclophosphamide workers (Hüttner *et al.*, 1990).

Several recent occupational studies using the autoradiographic *HPRT* assay have shown increases in VFs at exposure levels considered to be in the acceptable range. Three have involved exposures to butadiene. Thirteen non-smoking workers in a monomer production plant showed mean *HPRT* VFs of 1.0×10^{-6} for the six non-exposed controls, 1.2×10^{-6} for the five low-level exposed workers (0.03 ppm butadiene) and 4.0×10^{-6} for the eight high-level exposed workers (3.5 ppm butadiene). The mean VF of the highly exposed group was significantly elevated compared with the other groups. Furthermore, urinary levels of 1,2-dihydroxy-4- (*N*-acetylcysteinyl)butane, a marker of in-vivo butadiene dose, were significantly elevated in the highly exposed group and correlated significantly with *HPRT* VFs (Ward *et al.*, 1994). A second study of the same worker population and an ongoing study of workers in styrene-butadiene production tend to confirm these results (Ward *et al.*, 1996). Thus, these studies suggest that butadiene exposures at this level have genotoxic effects. (These results using the autoradiographic assay are remarkable because two studies with the cloning assay (in China and the Czech Republic; see below) have failed to show increases in *HPRT* MFs in populations with comparable butadiene exposures.)

The BrdU staining method of short-term assay has been used for only a single study of chemical exposure, i.e. to arsenic in Mexico (Ostrosky-Wegman *et al.*, 1991). Although the *HPRT* VFs in a high-level exposure group were reportedly twice the levels in a low-level exposure group (mean 5.0 vs mean 2.4×10^{-6}), this difference was not statistically significant.

HPRT mutant frequency (cloning assay): radiation. Cloning assays have been used to study *HPRT* mutations in Japanese atomic bomb survivors. All studies have shown low-level MF elevations in exposed individuals, with a 'shallow' dose–response curve (Hakoda *et al.*, 1988; Akiyama *et al.*, 1990; Hirai *et al.*, 1995). This was initially interpreted to indicate that *HPRT* is relatively insensitive to ionizing radiation. It is now realized, however, that the mutational signal has decayed in the 40+ years between the exposure and testing, as a consequence of mutations arising in peripheral T-lymphocytes without renewal from the bone marrow stem cell compartment in adults. In survivors who were young at the time of the exposures, the *HPRT* mutations also arose in precursor bone marrow stem cells—even in multipotent stem cells. This is probably the reason for any elevation at all in mean *HPRT* MF in the exposed populations. Indeed, a molecular study in one survivor showed that at least one mutational event arose in a stem cell before differentiation into the B-, NK and T-cell lineages had occurred (Hakoda *et al.*, 1989).

Several groups have found increased *HPRT* mutational responses to radiotherapy in cancer patients. From an early study of 12 breast cancer patients receiving 2 Gy/day local irradiation sampled shortly after treatment, it was estimated that T-cells circulating through the radiation field would have received 4 Gy irradiation, giving a radiation-induced increase of 6.9×10^{-6} mutants/Gy (Messing & Bradley, 1985). Others have found excesses for patients with breast cancer and Hodgkin's disease, although some investigators report only weak associations (Nicklas *et al.*, 1990, 1991; Sala-Trepat *et al.*, 1990; Branda *et al.*, 1991).

There have also been measurements of *HPRT* T-cell mutations in patients receiving extremely low irradiations (10–15 mGy) during nuclear medicine scans. Initial studies reported significantly increased mean MFs following these low exposure procedures (Seifert *et al.*, 1987), while later investigations did not find such an effect (Bachand *et al.*, 1991; Kelsey *et al.*, 1991; Van Dam *et al.*, 1991). It seems fair to conclude that the extremely low irradiations encountered in nuclear medicine scans are not measurable as mutagenic to *HPRT* and that the differences among studies were due to technical variables.

In contrast to patients, two early studies of radiation therapy and nuclear medicine technicians showed them to have elevated MFs relative to hospital controls based on doses received (2 mSv) in the previous 6 months (Messing *et al.*, 1986, 1989). This same investigative group has recently studied

workers exposed to low-level chronic irradiation (mean 1.34 mSv) in a Quebec factory (Seifert *et al.*, 1993). Although no group differences were found between exposed and non-exposed workers, radiation doses received 48–68 weeks before sampling were significantly positively correlated with MFs. The authors calculated a rate for induction of mutants of 0.7–3.4 $\times$ 10^{-4}/Sv, a rate which is similar to, but somewhat less than, that determined earlier for the radiation therapy and nuclear medicine technicians.

Of note, however, is a study of 36 workers at the Sellafield nuclear reprocessing installation, 18 of whom had cumulative recorded radiation doses of <50 mSv and 18 had cumulative doses of 500 mSv accumulated over many years, but which was clearly negative for *HPRT* mutations (Cole *et al.*, 1995).

Chemotherapy-related exposures. Many groups have used the cloning assay to measure *HPRT* MFs in adult patients receiving chemotherapy (Dempsey *et al.*, 1985; Palmer *et al.*, 1988; Sala-Trepat *et al.*, 1990; Branda *et al.*, 1991; Caggana *et al.*, 1991). Increased MFs were found among a group of patients with solid tumours and lymphomas, among breast cancer patients, other malignancies, and among patients treated with low doses of cyclophosphamide for connective tissue disease.

A recent study examined *HPRT* MFs in 15 cancer patients (10 with testicular cancer) and found that cyclophosphamide and ifosfamide were the most mutagenic agents, while adriamycin, 4-epi-adriamycin and bleomycin produced equivocal responses (Tates *et al.*, 1994a). Cisplatin and the etoposide VP16 (a topoisomerase inhibitor) did not increase *HPRT* mutations. These results are in accord with another study of multiple biomarker responses in germ cell tumour patients treated with platinum-based chemotherapeutic regimens, which showed marginal elevation in *HPRT* MFs (but increased *GPA* VFs) (Perera *et al.*, 1992a).

The mutagenicity of chemotherapy in children has been evaluated by the *HPRT* cloning assay. A study of 45 children with acute lymphoblastic leukaemia (ALL), 13 children with acute myelogenous leukaemia (AML) and 28 age-matched healthy controls showed a significantly higher mean MF (7.8×10^{-6}) in the treated ALL patients than in the treated AML patients (1.7×10^{-6}) and in healthy controls (1.1×10^{-6}) (Hirota *et al.*, 1993). Fifteen of the ALL patients had MFs >10×10^{-6} and elevations persisted for years. A potentially significant difference between the *HPRT* mutagenic responses induced by chemotherapy in adults and in children is the usually transient nature of the former and the persistence of the latter. This may indicate that, in children, a relatively larger proportion of the mutations are induced in bone marrow stem cells.

Other exposures. Oncology workers, nurses and pharmacists have all shown elevated *HPRT* MFs in different studies, although cloning efficiencies and lymphocyte subpopulations have also occasionally been affected by the exposures, which may have artefactually elevated values (Chrysostomou *et al.*, 1984; Dubeau *et al.*, 1994). Factory workers exposed to nor-nitrogen mustard clearly showed increased *HPRT* MFs compared with controls (Cole & Skopek, 1994).

Two groups of workers chronically exposed to ethylene oxide, i.e. nine hospital workers and 15 factory workers (40 h time-weighted average exposures of 0.25 ppm for the hospital workers and 5 ppm for the factory workers) showed mean MFs for the hospital workers of 12.4×10^{-6} and for the factory workers of 13.8×10^{-6} (Tates *et al.*, 1991). The latter elevation was significant compared with controls. The sensitivities of the various biomarkers employed in this study for detecting ethylene oxide exposure were haemoglobin adducts > sister chromatid exchanges > chromosome aberrations > micronuclei > *HPRT* mutations. By contrast, acute high-level exposures to ethylene oxide in seven workers had no effect on *HPRT*, SCE or any other biomarker evaluated, leading the authors to conclude that transient exposures to this agent, even at high doses, produce no genotoxic consequences (Tates *et al.*, 1995). In styrene workers, two *HPRT* studies have been negative and one suggestive of an effect in five subjects (Cole *et al.*, 1989; Tates *et al.*, 1994b; Vodicka *et al.*, 1995). In iron foundry factory workers exposed to PAHs at low levels, i.e. from <50 to 200 ng/m^3, *HPRT* mutations did correlate with borderline significance with estimated PAH exposure, but much more significantly with PAH adduct levels (Perera *et al.*, 1993, 1994). (By contrast, *GPA* mutations in this study were not sig-

nificantly elevated and did not correlate with DNA adducts.) In a study of bus maintenance workers, both the *HPRT* MFs and the PAH adduct levels were highest in the most heavily exposed workers, with a highly significant increase in individual MFs observed with increasing adduct levels (Hou *et al.*, 1995b). Two recent studies that used the cloning assay to measure *HPRT* MFs in butadiene-exposed workers have been negative (Hayes *et al.*, 1996; Tates *et al.*, 1996). These butadiene studies contrast with those mentioned above which used the autoradiographic assay.

In summary, ionizing radiation and chemicals at exposure levels found in accidental, household and occupational settings do induce *HPRT* mutational responses. The autoradiographic form of the assay may detect exposures with greater sensitivity than the cloning assay, but this must be demonstrated for the same populations receiving the same exposures for any valid conclusion.

HLA mutant frequency. The cloning assay for *HLA* mutations has not been used, thus far, to study humans exposed to mutagens.

HLA variant frequency (cytometric assay). The cytometric assay for *HLA* mutations was used to study 69 atomic bomb survivors, chosen as low-dose exposed (DS 86 doses = 0 Gy; 1986 estimates) or high-dose exposed (DS 86 doses >1 Gy) (Kushiro *et al.*, 1992). There were no significant increases in VFs with radiation. It was concluded from this study that *HLA* mutations cannot be detected in T-lymphocytes after the 40–50 years' lag between exposure and assay.

TCR variant frequency (cytometric assay). This assay has been used only to study radiation exposures. No significant dose effects were found in a large study of atomic bomb survivors, indicating again that mutational signals in T-cells have decayed in the 40+ years between exposure and test (Kyoizumi *et al.*, 1992). A single individual exposed to 3–4 Gy in Chernobyl was found to have an elevated VF of 21.1×10^{-4} 3.5 years later.

Eighteen thyroid cancer patients treated with ^{131}I from 2 months to 5 years earlier showed a significant linear relationship between the amount of radioactivity administered and the TCR gene VFs (Kyoizumi *et al.*, 1992). The induced mutations per administered dose were calculated as 0.30×10^{-4} TCR variants/GBq. Similarly, TCR VFs in six of 10 patients who received ^{232}Th for radiographic visualizations were elevated compared with concurrent controls (Umeki *et al.*, 1991; Kyoizumi *et al.*, 1992). ^{232}Th remains in the body and produces constant irradiation. By contrast, *GPA* VFs also measured in these patients showed no significant elevations.

Mutation spectra in reporter genes

The T-cell cloning assays allow for molecular analyses of mutations. By far the largest database is for *HPRT*, where mutation events ranging from single base changes to deletions, translocations and recombinations have been identified, and mutational spectra under a variety of circumstances have been described. *HLA* molecular analyses have been limited to Southern blots which have revealed a high frequency of events such as somatic recombination that involve the homologous chromosome. In contrast to autosomal genes, the X-chromosomal *HPRT* gene cannot undergo homologous recombination. However, the mutation spectra defined to date have been for *HPRT* mutations.

HPRT mutational spectra

Thousands of 'spontaneous' *HPRT* mutations arising *in vivo* in human T-cells have now been analysed at the molecular level. There have been several recent reviews of molecular studies, and a computerized database of published results is available (Albertini *et al.*, 1990; Cariello *et al.*, 1992; Cariello & Skopek, 1993; Cole & Skopek, 1994; Cariello, 1994).

The adult background *HPRT* mutational spectrum differs from that in the fetus and in young children. In adults, <15% of mutations arising *in vivo* show gross structural alterations such as deletions, insertions or other rearrangements on Southern blots (Nicklas *et al.*, 1989) (these Southern blot alterations involve >300 base pairs). The remaining 85% of adult *HPRT* mutations have been classified as 'point mutations' and include base substitutions, frame-shifts, smaller deletions and insertions, complex alterations and uncharacterized splice site changes. By contrast, the background *HPRT* mutations in placental cord blood show 75–85% to have gross structural alterations and the remainder to be 'point mutations' (Finette

et al., 1996), although maternal lifestyle factors and exposures may alter this pattern (Manchester *et al.*, 1995). Approximately 35% of fetal mutations show a single kind of DNA gross structural alteration, as discussed below. This reversal of the adult pattern persists in children until approximately the age of five.

The impetus for characterizing background mutation spectra in reporter genes is for comparison with mutational spectra following exposures to specific environmental mutagens. The expectation is that specific mutagens or classes of mutagens will induce characteristic mutational changes. Once identified, these changes can then be used in subsequent population studies to define the nature of exposures. In this sense, it is hoped that reporter gene mutations may serve as restricted biomarkers of exposure by providing specificity for the offending mutagen.

The discovery of induced in-vivo mutational spectra is just beginning. Ionizing radiation produces an *HPRT* mutational spectrum that becomes increasingly dominated by large structural alterations (such as deletions) as radiation doses increase, i.e. ionizing radiation, at least low energy transfer (LET) ionizing radiation, produces deletions at *HPRT* (Nicklas *et al.*, 1990, 1991). Studies of chemical mutagen exposures in humans have thus far given mixed results, probably because of insufficient numbers of mutants analysed. An early report indicated that in-vivo exposures to ethylene oxide induced a G to A transition at G_{197} (Cariello *et al.*, 1992). In-vitro controlled mutagenicity experiments are beginning to show characteristic mutational spectra, e.g. the pesticide malathion may be associated with characteristic deletions in exon 3 (Pluth *et al.*, 1996). Animal studies also show characteristic *HPRT* mutation spectra associated with particular mutagens, e.g. butadiene exposures in mice (Cochrane & Skopek, 1994). Therefore, some degree of mutational specificity from different mutagen exposures has been discovered at *HPRT*.

HLA mutational spectra. (Table 7) Thus far, the *HLA* mutations in human T-cells have been analysed only by Southern blots (Turner *et al.*, 1988; Morley *et al.*, 1990; Grist *et al.*, 1992). However, the investigations have defined loss of the specific target *HLA-A* gene in the mutants and the presence or absence of the *HLA-B* and other linked genes. Approximately 65% of background *HLA* mutants show no change on Southern blots, 2–8% show simple deletions, and 30% show changes compatible with mitotic recombination. Few show gene conversion. The 'no change' and mitotic recombination classes increase significantly with age.

In contrast to background, 75% of the *HLA* mutants induced by ionizing radiation *in vitro* show changes compatible with deletion (Kushiro *et al.*, 1992). Therefore, ionizing radiation is also characterized by deletion mutations at *HLA*.

Cancer-relevant mutational changes 'captured' in reporter genes (Table 3)

The rationale for defining *in vivo* reporter gene mutation spectra was originally to provide 'specificity' for identifying exposures. However, when used as biomarkers of effect, reporter gene mutations must also reflect events with pathogenic significance occurring elsewhere in the genome. To be useful as surrogates for cancer genes, reporter genes must undergo mutagenic processes that have carcinogenic potential.

Both the *GPA* and the *HLA* systems reflect mitotic recombinations among the mutants. This

Table 3. Cancer-relevant mutational changes: 'captured' in reporter genes

GPA	*HLA*	*HPRT*
Mitotic recombination	Mitotic recombination	Large deletions with topoisomerase II breakpoints
Gene conversion	Gene conversion	Fusion genes
Chromosome reduplication	Chromosome reduplication	V(D)J recombinase-mediated recombinations

mutational mechanism is known to underlie loss of heterozygosity (LOH) in several tumour suppressor genes, and, therefore, is an important mutational step in carcinogenesis. Agents that cause such changes in reporter genes, e.g. benzene-induced 'homozygous' *GPA* mutations as noted above, may be human carcinogens (Rothman *et al.*, 1995). Monitoring for such specific mutational changes may be important for predicting individual cancer outcomes.

Although the *HPRT* gene cannot undergo homologous somatic recombination, it does capture a variety of other carcinogenic mutagenic mechanisms. Large deletions and translocations, common in human tumours, are frequent changes in *HPRT* mutations (Nicklas *et al.*, 1989), particularly those following ionizing radiation (Nicklas *et al.*, 1990, 1991). The breakpoints of deletions often occur in DNA sequences with high homology to topoisomerase II consensus cleavage sequences where similar breakpoint sites are seen in the leukemias (Rainville *et al.*, 1995). *HPRT* mutations also may produce fusion genes, another change frequently observed in cancer.

One specific mutational change in *HPRT* is particularly striking because it mimics so well an event seen in virtually all lymphoid malignancies (Finger *et al.*, 1986; Boehm & Rabbitts, 1989; Tycko & Sklar, 1990; Breit *et al.*, 1993). This is the specific intragenic deletion that occurs *in vivo* during fetal life and early childhood, which is the most frequent single class of background *HPRT* mutations during this period of life (Fuscoe *et al.*, 1991; Manchester *et al.*, 1995; Finette *et al.*, 1996). These mutations show all of the characteristics of the V(D)J-mediated recombination that characterizes the TCR gene rearrangements, and they are virtually identical, at the sequence level, to the known cancer-related V(D)J recombinase-mediated mutations in lymphoid malignancies. This mutagenic mechanism with carcinogenic potential is precisely captured in *HPRT*.

As specific mutational mechanisms of carcinogenic significance are recognized, it becomes feasible to develop PCR and other molecular techniques for their rapid identification. Monitoring for these precise events, rather than reporter gene mutations in general, may be the relevant measures in reporter genes when they are used as biomarkers of effect in cancer epidemiology.

Somatic mutations in cancer genes

Several assays have been developed to detect gene mutations in cancer genes or their products.

Cancer-associated 'mutation' assays (Box 1)

Phenotypic 'mutation' assays. Malignant transformations are associated with mutations of oncogenes and tumour suppressor genes. In the case of the former, mutations result in overexpression of a normal protein (or expression of an aberrant protein) essential for cell proliferation. Many oncoproteins and tumour suppressor gene proteins are detectable by immunological techniques, e.g. immunoblot, ELISA, in body fluids such as serum from cancer patients (Brandt-Rauf, 1991, 1992). Studies of banked sera have revealed that, in many patients, the increased oncoprotein levels were present months to years before diagnosis. For environmentally related malignancies, it has been postulated that the causative carcinogen mutated the relevant gene(s) before the onset of clinical cancer. Measurements of these proteins might therefore be used as biomarkers of effect in epidemiological studies to detect cancer-relevant somatic mutations. Although many oncoproteins and tumour suppressor proteins have been studied, attention has focused on the *ras* p21 protein, the extracellular domain (ECD) of the c-*erbB-2* (HER-2, *neu*) p185 protein, the β-transforming growth factor (β-TGF) protein and the p53 protein.

Genotypic 'mutation' assays. Two polymerase chain reaction (PCR)-based molecular assays have been

Box 1. Cancer-associated 'mutation' assays

Phenotypic assays

- Oncogene protein in serum
  - *ras* p21
  - ECD of c-*erb* B-2 p185
  - β-TGF
- Tumour suppressor proteins in serum
  - p53

Genotypic assays

- Hybrid TCR genes
- BCL-2 rearrangements

introduced to detect 'mutations' associated with cancer. Strictly speaking, both measure forms of chromosome aberrations mediated by aberrant V(D)J recombinase activity, as described above. These two assays are described here because they measure directly either a cancer-relevant mutation mechanism or a known cancer-related chromosome translocation.

The first of these assays detects hybrid TCR genes, i.e. Vβ -Jγ or Vγ -Jβ, formed by inversions of chromosome 7 (inv7[p13; q32]) (Lipkowitz *et al.*, 1992). This inversion joins the TCRγ gene on chromosome 7p13-15 to the TCRβ gene on 7q32-35. (This molecular assay may measure events similar to the cytometric TCR gene somatic mutation assay described above, but direct comparisons have not been made.) By chosing primers for PCR from appropriate β and γTCR gene regions, products are formed only when the inversion has occurred. The assay is performed on peripheral blood lymphocytes, presumably the T-cells.

The second molecular method (BCL-2 translocation assay) measures an aberrant V(D)J recombinase-mediated event of direct relevance to cancer (Liu *et al.*, 1994). The chromosome translocation t(14:18) (q32:q21), which occurs at high frequency in non-Hodgkin's lymphomas, brings together the B-cell leukemia/lymphoma-2 (BCL-2) locus on chromosome 18 and the Ig heavy chain joining (J) region on chromosome 14, dysregulating the former and resulting in delayed programmed cell death. Again, by choosing appropriate primers for PCR, products are formed only when the translocation has occurred. The assay is performed on normal peripheral blood lymphocytes, presumably the B-cells.

Cancer-associated gene mutations in humans

The phenotypic assays. Immunological detection of oncogene and tumour suppressor gene products in sera was initially carried out in cancer patients to define the percentage of positives for the different cancers for the different products. As testing moved to serum banks, it became possible to assess sera retrospectively to determine if significant elevations of the serum markers could be found before the onset of cancer. Thus, the utility of these assays was originally for early diagnosis.

Oncogene and tumour suppressor gene proteins have now been used as biomarkers of effect in several human studies. Depending on the methods used and the products studied, these molecules may be found at low levels in healthy control individuals. Therefore, different studies have used different definitions of 'positive'. Furthermore, although some studies measured truly mutant proteins, most have simply measured increased levels of the marker protein, using methods that could not distinguish between mutant and normal molecules. The assumption is that elevated levels of even normal proteins are probably due to somatic mutations and correlate with cancer.

An early study of the *ras* p21 protein in healthy Finnish foundry workers exposed to PAHs in the workplace showed that one of the eight exposed and none of 10 unexposed individuals had detectable serum levels (Brandt-Rauf, 1992). In another study, three of 16 hazardous waste workers exposed to a wide variety of mutagens showed detectable serum *ras* p21 proteins, as did two of 17 unexposed workers (Brandt-Rauf, 1992). Both of the latter, however, were heavy smokers and were only 'trace' positive. One of the exposed workers developed a premalignant colon lesion 1.5 years after testing. Removal of this lesion normalized the serum *ras* p21 protein level.

Serum *ras* p21 proteins were determined in a Polish population study where a doubling in the frequency of high *ras* oncogene expression (>2SD control serum levels) was found in individuals exposed to environmental pollutants (Perera *et al.*, 1992b). A more recent study measured the *ras* p21 serum proteins in butadiene-exposed workers and found no elevations (Anderson *et al.*, 1996).

A study in vinyl chloride (VC) workers monitored the specific mutant Asp13c-Ki-*ras* p21 protein (DeVivo *et al.*, 1994). Four of five exposed workers with liver angiosarcoma and eight of nine with liver angiomas had detectable mutant proteins (by immunoblotting) in their sera. Importantly, 22 of 45 (49%) of the VC-exposed workers with no evidence of liver neoplasia also showed detectable levels. A significant linear trend was found for mutant p21 protein in serum and increasing duration of VC exposure. None of 28 non-exposed individuals had detectable serum mutant p21 protein.

Other oncoproteins have also been studied in mutagen/carcinogen-exposed individuals. Forty-six pneumoconiosis patients (32 asbestos, 10 silicosis) were studied for a variety of oncoproteins

(Brandt-Rauf *et al.*, 1992). Five of 18 with cancer had elevated serum *ras* p21 proteins (total), defined as a fivefold elevation over normal by dilution, compared with only two of the 28 without cancer. There were many pre-diagnosis positive serum values in this study. Pneumoconiosis patients also had significant elevations of platelet-derived growth factor (PDGF) serum proteins.

Another study of multiple serum oncoproteins showed that the β-TGF proteins were elevated in 14 of 33 fire-fighters exposed to a variety of pulmonary mutagens/carcinogens (Ford *et al.*, 1992). No elevations were found in unexposed controls.

Studies are being reported of serum levels of the extracellular domain (ECD) of the c-*erbB-2* protein p185 and the epidermal growth factor receptor (EGF) protein (Brandt-Rauf *et al.*, 1994; Partanen *et al.*, 1994). These are elevated in certain premalignant conditions, in early cancers, and in some exposure situations with a high risk of cancer. Studies of p53 serum protein in early cancers are ongoing. All are being pursued with the goal of developing biomarkers of effect for human mutagenicity monitoring.

The history of the of cancer gene biomarkers is that the original study populations were individuals with cancers. Studies then progressed to individuals with early cancers, then to individuals with premalignant conditions, and finally to individuals exposed to mutagens/carcinogens with a high risk of cancer. This contrasts with the history of the reporter gene mutation assays. In this case, development progressed from studies of healthy individuals exposed to mutagenic agents and were designed to detect exposures. Only recently has attention focused on cancer-relevant mutagenic mechanisms. Thus, development of these two kinds of assays have progressed from opposite directions towards each other, i.e. reporter genes moving from exposure to disease, with the cancer gene mutations moving from disease to exposures.

The genotypic assays. The PCR-based method for detecting the inv7 (p13:q32) chromosome aberration has been used to study normal individuals, patients with AT, and individuals exposed to pesticides (Lipkowitz *et al.*, 1992). Normal background frequencies are 10^{-5}, as determined by limiting dilutions. AT homozygotes have a 100-fold increase in frequency, while heterozygous carriers have normal values. Importantly, individuals heavily exposed to pesticides have had inversion frequencies intermediate between normal control and AT patients.

Standardization of the PCR method for detecting BCL-2 translocation frequencies is just beginning (Liu *et al.*, 1994). Quantification is achieved by a multiple tube method based on Poisson distributions. Individuals with no detectable translocations ($<10^{-6}$) are considered as 'negative.' Twenty-four of 53 blood samples from normal individuals were negative for this translocation. Of the positives, translocation frequencies varied from 0.8 to 32.0×10^{-6}, a 40-fold difference between the lowest and the highest (values $<10^{-6}$ can be detected when frequencies are based on several analyses). When stratified into age groupings of 0–20 years, 21–40 years, 41–60 years and >60 years, these frequencies were 0.29, 0.77, 1.43 and 3.39×10^{-6}, respectively. Variability also increased with age. (It should be noted that, although termed 'translocation frequencies', many of the translocations from specific individuals represent large clones, some of which have persisted for long intervals. Therefore, in these cases the measured frequencies greatly overestimate the translocation events.)

Issues in the application of somatic mutations in epidemiological studies

Strengths and limitations of the methods

At present, seven assays are available for assessing mutations of five reporter genes using two cell types. Mutations scored in RBCs occur in nucleated precursor cells, thus limiting the in-vivo site of mutation to the bone marrow. Moreover, the RBC assays, i.e. Hb and *GPA*, do not allow for molecular analyses of the mutations or the development of direct molecular detection methods. The advantages of the RBC assays is that they require very small blood samples and are rapid, inexpensive and simple. Furthermore, the *GPA* assay, although phenotypic, allows for assessment of mitotic recombination, an important mechanism in LOH. Finally, when mutations arise in the multipotent bone marrow stem cells, they are potentially long-lived and may be used for remote dose reconstructions or nested case–control studies. (However, chemically induced mutations in RBC precursors appear to arise, in part, in more differentiated cells with much shorter persistence, i.e. only months.)

The greatest strength of the lymphocyte assays is that they provide nucleated cells for molecular analyses. Mutation spectra can be determined and direct molecular assay methods developed. Furthermore, as mutations occur in peripheral T-cells, mutations can arise in all body sites. A limitation of T-cells, however, is that tissue culture methods must be used for molecular analyses to be possible. These are usually slow, costly and labour-intensive. In addition, relatively large blood samples must be obtained. The short-term lymphocyte assays circumvent this, using inexpensive and rapid methods, but do not allow for molecular analyses. The mutational memory in the peripheral T-cells is probably a matter of months, at least in adults. Therefore, these mutations will not be of value for remote past exposures.

The time of appearance of mutants in peripheral cells has not been precisely defined for all of the mutation assays. Mutants appear as early as 2 weeks after mutagen exposures, as shown for the *HPRT* autoradiographic assay (Ammenheuser *et al.*, 1988). Usually, however, the time for optimal occurrence will be a period of months. Additional longitudinal studies are needed to establish the time of optimum mutant appearance for various exposures. In any case, mutant cells will usually appear long after the metabolites, SCEs and chromosomal aberrations when these biomarkers are used with somatic mutations in epidemiological studies. The use of multiple biomarkers, however, is usually advantageous in epidemiological studies as some, i.e. DNA adducts, are the best estimators of in-vivo mutagen doses. Several studies have shown correlations of *HPRT* mutations to these biomarkers of exposure but not to ambient exposures.

Assay variability

All of the currently used reporter gene assays show large interindividual variation. Much of this is biological, and epidemiological studies will help to establish the precise causes. It has been shown that the rare genetic instability syndromes are associated with large increases in VFs and MFs. It is not known, however, if lesser deficiencies of DNA repair will also be reflected in detectable increases in mutations. This will require methods to assess more accurately the repair capabilities themselves. An important emerging area of research is the correlation between the various metabolic genotypes and the interindividual differences in somatic mutations. Since these genotypes are associated with different responses for the biomarkers of exposure, i.e. DNA adducts, there is every reason to expect that somatic mutations will also be increased in susceptible individuals. A suggested association between increased *HPRT* MF in individuals of GSTM1 null genotype has recently been reported (Hou *et al.*, 1995a). Eventually, epidemiological studies will have to take all of these into account in assessing somatic mutations.

Physiological, nutritional and pathological changes can also influence interindividual and even intra-individual replicate sample variations in somatic mutations. For example, recent studies have shown that *HPRT* mutations are inversely associated with serum folate levels (Branda *et al.*, 1991). Viral infections may increase the mutability of some somatic cells (Havre *et al.*, 1995). Autoimmune diseases raise in-vivo T-cell *HPRT* MFs, probably as a consequence of increased cell proliferation (Theocharis *et al.*, 1995). Haematopoietic stress may also affect RBC assays. Further studies will also have to relate VFs and MFs to these factors.

This variability obviously affects the ability to detect differences in VFs or MFs between groups in epidemiological studies. A systematic analysis of required sample sizes, given various coefficients of variation, has been accomplished for the *HPRT* cloning assay (Robinson *et al.*, 1994). Similar analyses must be undertaken for all of the systems.

Superimposed on the biological variability is the technical variability. Cryo- or other standards must be developed and validated over time for all of the assays used in epidemiological studies.

Somatic mutations as biomarkers of exposure

Exposure/dose assessment. Somatic mutations in reporter genes are used to assess mutagen/carcinogen exposures. A major determinant of their utility in this regard is their sensitivity relative to other biological end-points. Although mutations do detect exposures, it has been shown that, compared to other measures, e.g. metabolites or protein or DNA adducts, they are among the least sensitive for chemical mutagens/carcinogens (Tates *et al.*, 1991). This might have been expected as the metabolites and adducts are more proximal to the exposures. However, even for acute ionizing irradiation, the most sensitive of the somatic mutation assays is no

better and probably less sensitive than are chromosome aberrations. The reason is that the latter, although clearly reflecting a genotoxic effect, is a very large target, i.e. the entire genome.

Although lacking in sensitivity, there are occasions when reporter gene mutations may be used as exposure dosimeters. First, those assays that are simple and require only small blood samples may be easier and less expensive to use than other biomarkers. Moreover, mutations may be used for exposure assessment for unknown mutagens/carcinogens when no other biological end-points are available. Another use for certain somatic mutations for exposure/dose assessment may be for dose reconstructions of remote exposures, as in nested case–control studies. However, only mutation assays that detect events in stem cells, i.e. the RBC assays, will be useful in this regard. Finally, somatic mutations in the fetus, i.e. as *HPRT* in T-cells, may be useful for population exposure assessments because the newly induced mutations arise on a low and characteristic background.

Exposure characterization. Somatic mutations as biomarkers of exposure may have their greatest utility in terms of exposure characterizations. Mutation spectra are being defined with the expectation that the naturally occurring background spectrum will differ from those produced by different mutagens or classes of mutagens. If so, characterizing the mutations will allow a diagnosis of a specific exposure, i.e. will provide specificity. This will be possible only for those mutations that are recovered for molecular analyses—currently requiring the T-cell assays and laborious tissue culture methods.

It has been shown that ionizing radiation produces a characteristic spectrum of deletions for both *HPRT* and *HLA* mutations. A good deal of effort is now being expended in defining chemical molecular mutational spectra using both in-vivo and in-vitro systems. The identification of such spectra for exposure characterizations may become an important reason for mutagenicity monitoring in humans.

Somatic mutations for detecting genotoxic effects

Somatic mutations unequivocally reveal genotoxic effects. Although such effects in reporter genes do not necessarily indicate genotoxic effects in cancer genes, the most important potential application of reporter gene mutations for human biomonitoring may eventually be as surrogates for cancer mutations. For this to become a reality, it must be demonstrated that mutations in reporter genes, measured in tissues of convenience, are valid surrogates for mutations in cancer genes occurring in target tissues.

Current evidence that reporter gene mutations are valid surrogates in this regard is indirect but positive. First, there is the analogy to another biomarker of effect, namely non-specific chromosome aberrations. Two recent retrospective follow-up studies have shown that individuals with high frequencies of non-specific chromosome changes have relative risks of developing cancer in the next decade of greater than 2.0 (Hagmar *et al.*, 1994; Bonassi *et al.*, 1995). Regression analyses indicated that the aberrations conferred risk beyond that associated with exposure *per se* (Bonassi *et al.*, 1995). It is noteworthy that SCE frequencies, which often detect exposures more sensitively, were not associated with increases in cancer risk.

Animal studies have shown that agents that produce cancers in various tissues also produce *HPRT* T-cell mutations *in vivo* (Aidoo *et al.*, 1991). Thus, *HPRT* is a functional surrogate for cancer in these species. Finally, the administration of radioprotective agents in mice receiving ionizing radiation reduced both the induced malignancies and *HPRT* T-cell mutations, again relating the reporter events to disease-causing events (Grdina *et al.*, 1991, 1992).

The best evidence that reporter mutations reflect the occurrence of cancer gene mutations in humans is the discovery that mutagenic mechanisms with carcinogenic potential are captured in these reporter genes. The occurrences of the somatic recombinations, the deletions with characteristic breakpoints, the fusion genes and the V(D)J recombinase-mediated mutations in the various reporter genes have been described.

The use of the various assays detecting events in cancer genes for epidemiological studies remains to be defined.

Future directions

Next-generation assay development

The next generation of assay development will probably involve both reporter and cancer genes. Reporter genes are useful for defining the molecular

bases of in-vivo genetic damage. However, the assays that allow this are costly and difficult to perform. New developments will therefore be in methods to identify and quantify mutations that do not require tissue culture.

The five current reporter gene mutation assays use only blood cells. The development of molecular-based assays will also allow the sampling of other non-selectable genes and will include examination of other tissues. Eventually, a large armamentarium can be developed, using the tissue culture method to design and characterize systems and the molecular technologies for application to large-scale human studies.

The development of mutation assays for changes in cancer genes is just now beginning. Future assay development will likely focus on exposure-specific mutations or mutated protein products. Specific and sensitive immunological and molecular assays require further methodological development.

Increased understanding of the biological process of somatic mutation

The interpretation of somatic mutations in human populations will be enhanced by increased understanding of the biological basis of their occurrence and of how this relates to human carcinogenesis. Advances in our understanding of the biological basis for several of the major gene instability syndromes have led to an appreciation of the underlying mechanisms for the frequent occurrence of somatic mutations in these syndromes. Other mechanistic links may be expected, as indicated above for V(D)J recombinase. It has recently been demonstrated that transfection of human cells *in vitro* with HPV proteins may inactivate *p53* and increase mutagenesis of the *HPRT* reporter gene (Havre *et al.*, 1995). Thus, somatic mutations in human populations may be determined by infections as well as by chemical/ physical mutagen exposures and underlying genetic susceptibilities. Studies in human populations exploring the mechanisms for somatic cell mutation will contribute to an understanding of these interrelated factors, and consequently of a possible relation to cancer.

Although the 'memory' of somatic mutations appears to diminish with time, the temporal sequence of development, appearance in the peripheral blood (or other potentially assessed tissues), and eventual disappearance of somatic mutations are poorly understood. As our understanding of cell replication and death increases, it may be possible to distinguish more clearly long-term mutational events. This would clearly be advantageous for the purposes of dosimetry. This insight, however, may also be relevant for understanding the time-dependent relationship between exposure and the development of cancer. For example, radiation-induced leukemia is detectable after just a few years, reaches a peak at about 5 years after exposure, and decreases thereafter (Committee on the Biological Effects of Ionizing Radiation, 1990). A similar pattern, but with a substantially longer wave period, has been described for radon-induced lung cancer (Lubin *et al.*, 1994). Studies to evaluate the time-dependent occurrence of somatic mutations in cell populations may provide insight into this process.

With the further development of molecular approaches to somatic mutation assessment, it will increasingly be possible to investigate multiple markers in human studies. This approach can provide internal consistency checks and may broaden mechanistic understanding.

Validation of somatic mutations as predictors of disease outcome

A limitation for the interpretation of the reporter gene somatic cell mutation assays is uncertainty about the significance of these events as surrogates for mutations leading to cancer in humans. For mutations in cancer genes, problems in interpretation also arise because the mutations observed may not lead to functional alterations or may be 'after the fact', i.e. may indicate existing but subclinical cancer, and therefore not be useful for the purposes of prevention.

With improvements in assay methodology, it will increasingly be possible in large-scale epidemiological studies to determine if specific somatic cell mutations are statistically associated with increased risks for cancer. Case–control studies can be used for this purpose, but will probably be of limited value for the 'short memory' markers and because of the possibility that disease status may influence marker outcome. 'Nested' case–control studies of a subset of cases and comparison subjects for whom samples were collected prior to disease development as part of a large cohort investigation provide a methodologically sound alternative to

the case-control study. Several such cohorts have now been established, but procedures for the collection and storage of biological materials are suitable only for molecular studies and not for the standard cell-based assays, unless special procedures such as cryopreservation are undertaken. Although a case can be made for the special storage procedures, they are laborious and expensive. Because of this, direct validation studies will usually require molecular approaches.

While such validation studies are clearly needed to describe the statistical relationship between somatic mutations and cancer risk, a positive association does not establish a causal link. Ultimately, the usefulness of the somatic cell mutation assays will be based upon our understanding of their biological basis and of how this relates to human carcinogenesis.

References

Aldoo, A., Lyn-Cook, L.E., Mittelsyaedt, R.A., Heflich, R.H. & Casciano, D.A. (1991) Induction of 6-thioguanine-resistant lymphocytes in Fischer 344 rats following *in vivo* exposure to *N*-ethyl-*N*-nitrosourea and cyclophosphamide. *Environ. Mol. Mutagen.*, 17, 141–151

Akiyama, M., Kyoizumi, S., Hirai, Y., Hakoda, M., Nakamura, N. & Awa, A.A. (1990) Studies on chromosome aberrations and HPRT mutations in lymphocytes and GPA mutation in erythrocytes of atomic bomb survivors. In: Mendelson, M.L. & Albertini, R.J., eds, *Mutation and the Environment Part c: Somatic and Heritable Mutation, Adduction, and Epidemiology, Progress in Clinical and Biological Research, 340c*, New York, Wiley-Liss, pp. 69–80

Albertini, R.J. (1985a) Somatic gene mutations *in vivo* as indicated by the 6-thioguanine-resistant T-lymphocytes in human blood. *Mut. Res.*, 150, 411–422

Albertini, R.J. (1985b) The use of human T-lymphocytes to monitor mutations in human populations. In: Li, A.P., ed., *New Approaches in Toxicity Testing and Their Application in Human Risk Assessments*. New York, Raven Press. pp. 51–66

Albertini, R.J., Allen, E.F., Quinn, A.S. & Albertini, M.R. (1981) Human somatic cell mutation: *in vivo* variant lymphocyte frequencies determined by 6-thioguanine resistant lymphocytes. In: Hook, E.B. & Potter, J.H., eds, *Population and Biological Aspects of Human Mutation*, New York, Academic Press, pp. 235–263

Albertini, R.J., Castle, K.L. & Borcherding, W.R. (1982) T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood. *Proc. Natl Acad. Sci. USA*, 79, 6617–6621

Albertini, R.J., Sullivan, L.S., Berman, J.K., Greene, C.J., Stewart, J.A., Silveira, J.M. & O.'Neill, J.P. (1988) Mutagenicity monitoring in humans by autoradiographic assay for mutant T-lymphocytes. *Mut. Res.*, 204, 481–492

Albertini, R.J., Nicklas, J.A., O.'Neill, J.P. & Robison, S.H. (1990) *In vivo* somatic mutations in humans: measurement and analysis. *Ann. Rev. Genet.*, 24, 305–326

Ammenheuser, M.M., Ward J.B., Jr, Whorton, E.B., Jr, Killian, J.M. & Legator, M.S. (1988) Elevated frequencies of 6-thioguanine resistant lymphocytes in multiple sclerosis patients treated with cyclophosphamide: a prospective study. *Mut. Res.*, 204, 509–520

Ammenheuser, M.M., Au, W.W., Whorton E.B., Jr, Belli, J.A. & Ward J.B., Jr (1991) Comparison of *HPRT* variant frequencies and chromosome aberration frequencies in lymphocytes from radiotherapy and chemotherapy patients: a prospective study. *Environ. Mol. Mutagen.*, 18, 126–135

Anderson, D., Hughes, J.A., Brinkworth, M.H., Peltonen, K. & Sorsa, M. (1996) Levels of *ras* oncoproteins in human plasma from 1,3-butadiene-exposed workers and controls. *Mut. Res.*, 349, 115–120

Bachand, M., Seifert, A.M. & Messing, K. (1991) Nuclear medicine patients do not have higher mutant frequencies after exposure thallium-201. *Mut. Res.*, 262, 1–6

Bernini, L.F., Natarajan, A.T., Schreuder-Rotteveel, A.H.M., Giordano, P.C., Ploem, J.S. & Tates, A. (1990) Assay for somatic mutation of human hemoglobins. In: Mendelson, M.L. & Albertini, R.J., eds, *Mutation and the Environment Part c: Somatic and Heritable Mutation, Adduction, and Epidemiology, Vol Progress in Clinical and Biological Research, 340c*, New York, Wiley-Liss, pp. 57–68

Bigbee, W.L., Branscombe, E.W., Weintraub, H.B., Papayannapoulou, T. & Stamatoyannopoulos, G. (1981) Cell sorter immunofluorescence detection of human erythrocytes labeled in suspension with antibodies specific for hemoglobin S and C. *J. Immunol. Methods*, 45, 117–127

Bigbee, W.L., Wryobek, R.G., Langlois, R.G., Jensen, R.H. & Everson, R.B. (1990) The effect of chemotherapy on the in vivo frequency of glycophorin A "null" variant erythrocytes. *Mut. Res.*, 240, 165–175

Bigbee, W.L., Grant, S.G., Langlois, R.G., Jensen, R.H., Anttila, A., Pfaffli, P., Pekari, K. & Norppa, H. (1996) Glycophorin A (*GPA*) somatic cell mutation frequencies in Finnish reinforced plastics workers exposed to styrene. *Cancer Epidemiol. Biomarkers Prev.*, 5, 801–810

Bodmer, W.F. (1984) The HLA system, 1984. In: Albert, E.D., Baur, M.P. & Mayr, W.R., eds, *Histocompatibility Testing 1984*, Berlin, Springer Verlag, pp. 11–22

Boehm, T. & Rabbitts, T.H. (1989) The human T cell receptor genes are targets for chromosomal abnormali-

ties in T cell tumors. *FASEB J*, 3, 2344–2359

Bonassi, S., Abbondandolo, A., Camurri, L., Dal Prá, L., De Ferrari, M., Degrassi, F., Forni, A., Lamberti, L., Lando, C., Padovani, P., Sbrana, I., Vecchio, D. & Puntoni, R. (1995) Are chromosome aberrations in circulating lymphocytes predictive of future cancer onset in humans? Preliminary results of an Italian cohort study. *Cancer Genet. Cytogenet.*, 79, 133–135

Branda, R.F., O.'Neill, J.P., Sullivan, L.M. & Albertini, R.J. (1991) Factors influencing mutation at the *HPRT* locus in T-lymphocytes: Women treated for breast cancer. *Cancer Res.*, 51, 6603–6607

Brandt-Rauf, P.W. (1991) Oncogene proteins as biomarkers in the molecular epidemiology of occupational carcinogenesis, The example of the *ras* oncogene-encoded p21 protein. *Intl Arch. Occup. Environ. Health*, 63, 1–8

Brandt-Rauf, P.W. (1992) Oncogenes and oncoproteins in occupational carcinogenesis. *Scand. J. Work Environ. Health*, 18, 27–30

Brandt-Rauf, P.W., Smith, S., Hemminki, K., Koskinen, H., Vainio, H., Niman, H. & Ford, J. (1992) Serum oncoproteins and growth factors in asbestosis and silicosis patients. *Int. J. Cancer*, 50, 881–885

Brandt-Rauf, P.W., Pincus, M.R. & Carney, W.P. (1994) The c-*erbB*-2 protein in oncogenesis: molecular structure to molecular epidemiology. *Crit. Rev. Oncogen*, 5, 313–329

Breit, T.M., Mol, E.J., Wolvers-Tettero, I.L.M., Ludwig W.-D, van Wering, E.R. & Van Dongen, J.J.M. (1993) Site-specified deletions involving the *tal*-1and *sil* genes are restricted to cells of the T-cell receptor lineage: T-cell receptor gene deletion mechanism affect multiple genes. *J. Exp. Med.*, 177, 965–977

Caggana, M., Liber, H.L., Mauch, P.M., Coleman, C.N. & Kelsey, K.T. (1991) *In vivo* somatic mutation in the lymphocytes of Hodgkin's disease patients. *Environ. Mol. Mutagen.*, 18, 6–13

Cariello, N.F. (1994) Software for the analysis of mutations at the human *HPRT* gene. *Mut. Res.*, 312, 173–185

Cariello, N.F., Craft, T.R., Vrieling, H., van Zeeland, A.A., Adams, T. & Skopek, T.R. (1992) Human HPRT mutant database: software for data entry and retrieval. *Environ. Mol. Mutagen.*, 20, 81–83

Cariello, N.F. & Skopek, T.R. (1993) Analysis of mutation occurring at the human *HPRT* locus. *J. Mol. Biol.*, 231, 41–57

Cartron J.-P, Collin, Y., Kudo, S. & Fukuda, M. (1990) Molecular genetics of human erythrocyte sialoglycoproteins: Glycophorins A, B, C and D. In: Harris, J.R., ed., *Blood Cell Biochemistry* (Vol. 1). New York, Plenum, pp. 299–335

Chrysostomou, A., Seshadri, R.S. & Morley, A.A. (1984) Mutation frequency in nurses and pharmacists exposed to cytotoxic drugs. *Aus. NZ J. Med.*, 14, 831–834

Clevers, H.B., Alarcon, T., Wileman, C. & Terhost, C. (1988) The T cell receptor/CD3 complex: a dynamic protein ensemble. *Ann. Rev. Immunol.*, 6, 629–662

Cochrane J.E. & Skopek, T.R. (1994) Mutagenicity of butadiene and its epoxide metabolites: II. Mutational spectra of butadiene, 1,2-epoxybutene and diepoxybutane at the *HPRT* locus in splenic T cells from exposed B6C3F1 mice. *Carcinogenesis*, 15, 719–723

Cole, J. & Skopek, T.R. (1994) International Commission for Protection against Enviormental Mutagens and Carcinogens, Working paper 3, Somatic mutant frequency, mutation rates and mutational spectra in the human population *in vivo*. *Mut. Res.*, 304, 33–106

Cole, J., Arlett, C.F., Green, M.H.L., James, S.E., Henderson, L., Cole, H., Sala-Trepat, M., Benzi, R., Price, M.L. & Bridges, B.A. (1989) Measurement of mutant frequency to 6-thioguanine resistance in circulating T-lymphocytes for human population monitoring. In: Jolles, G. & Cordier, A., eds, *New Trends in Genetic Risk Assessment*. London, Academic Press, pp. 175–203

Cole, J., Arlett, C.F., Green, M.H., Holdsworth, D., Tawn, E.J. & Bridges, B.A. (1995) Mutant frequencies in workers at the Sellafield installation. *Health Phys.*, 68, 388–393

Committee on the Biological Effects of Ionizing Radiation (1990) *Health Effects of Exposure to Low Levels of Ionizing Radiation*, B.E.I.R V. Washington, DC, National Academy Press, pp. 1–421

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W.F., Ryddle, F., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977) Localization of the human alpha-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell*, 12, 205–218

Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. & Ruddle, F.H. (1978) Chromosomal localization of human beta-globin gene on human chromosome 11 in somatic cell hybrids. *Proc. Natl Acad. Sci. USA*, 75, 1456–1460

Dempsey, J.L., Seshadri, R.S. & Morley, A.A. (1985) Increased mutation frequency following treatment with cancer chemotherapy. *Cancer Res.*, 45, 2873–2877

DeVivo, I., Marion M.-J, Smith, S.J., Carney, W.P. & Brandt-Rauf, P.W. (1994) Mutant c-Ki-*ras* p21 protein in chemical carcinogenesis in humans exposed to vinyl chloride. *Cancer Causes Control*, 5, 273–278

Dubeau, H., Zazi, W., Baron, C. & Messing, K. (1994)

Effects of lymphocyte subpopulations on the clonal assay of HPRT mutants: occupational exposure to cytostatic drugs. *Mut. Res.*, 321, 147–157

Finette, B.A., Poseno, T. & Albertini, R.J. (1996) V (D)J recombinase mediated HPRT mutations in peripheral blood lymphocytes of normal children. *Cancer Res.*, 56, 1405–1412

Finger, L.R., Harvey, R.C., Moore, R.C.A., Showe, L.C. & Croce, C.M. (1986) A common mechanism of chromosomal translocation in T- and B-cell neoplasia. *Science*, 234, 982–985

Ford, J., Smith, S., Luo J.-C, Friedman-Jimanez, G., Brandt-Rauf, P., Markowitz, S., Garibaldi, K. & Niman, H. (1992) Serum growth factors and oncoproteins in firefighters. *Occup. Med.*, 42, 39–42

Furthmayer, H. (1977) Structural analysis of a membrane glycoprotein: glycophorin A. *J. Supramol. Struct.*, 7, 121–134

Furthmayer, H. (1978) Structural comparison of glycophorins and immunochemical analysis of genetic variants. *Nature*, 271, 519–524

Fuscoe, J.C., Zimmerman, L.J., Lippert, J.J., Nicklas, J.A., O.'Neill, J.P. & Albertini, R.J. (1991) V (D)J recombinase-like activity mediates *HPRT* gene deletion in human fetal T-lymphocytes. *Cancer Res.*, 51, 6001–6005

Gahmberg, C.G., Jakinen, M. & Andersson, L.C. (1979) Expression of the major red cell sialoglycoprotein, glycophorin A, in the human leukemic cell line K562. *J. Biol. Chem.*, 254, 7442–7448

Grant, S. & Bigbee, W.L. (1993) *In vivo* somatic mutation and segregation at the human glycophorin A (GPA) locus. *Mut. Res.*, 288, 163–172

Grant S.G. & Bigbee, W.L. (1994) Bone marrow somatic mutation following genotoxic cancer therapy. *Lancet*, 343, 1507–1508

Grdina, D.J., Carnes, B.A., Grahn, D. & Sigdestad, C.P. (1991) Protection against late effects of radiation by S-2-(3-aminopropylamino)-ethylphosphorothioic acid. *Cancer Res.*, 51, 4125–4130

Grdina, D.J., Kataoka, Y., Basic, I. & Perrin, J. (1992) The radioprotector WR-2721 reduces neutron-induced mutations at the hypoxanthine-guanine phosphoribosyl transferase locus in mouse splenocytes when administered prior to or following irradiation. *Carcinogenesis*, 13, 811–814

Grist, S.A., McCarron, M., Kutlaca, A., Turner, D.R. & Morley, A.A. (1992) *In vivo* human somatic mutation: frequency and spectrum with age. *Mut. Res.*, 266, 189–196

Hagmar, L., Brogger, A., Hansteen, I.-L., Heim, S., Högstedt, B., Knudsen, L., Lambert, B., Linnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerfving, S. & Sorsa, M. (1994) Cancer risks in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic study group on the health risk of chromosome damage. *Cancer Res.*, 54, 2919–2922

Hakoda, M., Akiyama, M., Kyoizumi, S., Awa, A.A., Yamakido, M. & Otake, M. (1988) Increased somatic cell mutant frequency in atomic bomb survivors. *Mut. Res.*, 201, 39–48

Hakoda, M., Hirai, Y., Shimba, H. & Kusunoki, Y. (1989) Cloning of phenotypically different human lymphocytes originating from a single stem cell. *J. Exp. Med.*, 169, 1265–1276

Havre, P.A., Yuan, J., Hedrick, L., Cho, K.R. & Glazer, P.M. (1995) P53 inactivation by HPV16 E6 resuslts in increased mutagenesis in human cells. *Cancer Res.*earch, 55, 4420–4424

Hayes, R.B., Xi, L., Bechtold, W.E., Rothman, M., Yao, M., Henderson, R., Zhang, L., Smith, M.T., Zhang, D., Wiemels, J., Dosemeci, M., Yin, S. & O.'Neill, J.P. (1996) *HPRT* mutation frequency among workers exposed to 1,3-butadiene in China. *Toxicology*, 113, 100–105

Henderson, L., Cole, H., Cole, J., James, S.E. & Green, M. (1986) Detection of somatic mutations in man: evaluation of the microtitre cloning assay for T-lymphocytes. *Mutagenesis*, 1, 195–200

Hewitt, M. & Mott, M.G. (1992) The assessment of *in vivo* somatic mutations in survivors of childhood malignancy. *Br. J. Cancer*, 66, 143–147

Hirai, Y., Kusunoki, Y., Kyoizumi, S., Awa, A.A., Pawel, D.J., Nakamura, N. & Akiyama, M. (1995) Mutant frequency at the HPRT locus in peripheral blood T-lymphocytes of atomic bomb survivors. *Mut. Res.*, 329, 183–196

Hirota, H., Kubota, M., Hashimoto, H., Adachi, S., Matsubara, K., Kuwakado, K., Akiyama, Y., Tsutsui, T. & Mikawa, H. (1993) Analysis of *HPRT* gene mutation following anti-cancer treatment in pediatric patients with acute leukemia. *Mut. Res.*, 319, 113–120

Hou, S.-M., Falt, S. & Steen, A.-M. (1995a) Hprt mutant frequency and GSTM1 genotype in non-smoking healthy individuals. *Environ. Mol. Mutagen*, 25, 97–105

Hou, S.-M., Lambert, B. & Hemminki, K. (1995b) Relationship between *HPRT* mutant frequency, aromatic DNA adducts and genotypes for GSTM1 and NAT2 in bus maintenance workers. *Carcinogenesis* 16: 1913–1917

Hüttner, E., Mergner, U., Braun, R. & Schoneich, J. (1990) Increased frequency of 6-thioguanine-resistant lymphocytes in peripheral blood of workers employed in cyclophosphamide production. *Mut. Res.*, 243, 101–107

Janatipour, M., Trainor, K.J., Kutlaca, R., Bennett, G., Hay, J., Turner, D.R. & Morley, A.A. (1988) Mutations in

human lymphocytes studied by an HLA selection system. *Mut. Res.*, 198, 221–226

Janeway, C.A., Jr & Travers, P. (1996) Chapter 4: antigen recognition by T-lymphocytes. In: *Immunobiology: The Immune System in Health and Disease*, 2nd edn, London, Current Biology Ltd, pp. 38–48

Jensen, R.H. & Bigbee, W. (1996) Direct immunofluorescence labeling provides an improved method for the glycophorin A somatic cell mutation assay. *Cytometry*, 23, 337–343

Jensen, R.H., Langlois, R.G., Bigbee, W.L., Grant, S.G., Moore D., III, Pilinskaya, M., Vorobtsova, I. & Pleshanov, P. (1995) Elevated frequency of glycophorin A mutations in erythrocytes from Chernobyl accident victims. *Rad. Res.*, 141, 129–135

Kelsey, K.T., Donohoe, J.K., Memisoglu, A., Baxter, B., Caggana, M. & Liber, H.I. (1991) *In vivo* exposure of human lymphocytes to technitium-99m in nuclear medicine patient does not induce detectable genetic effects. *Mut. Res.*, 264, 213–218

Kudo, S. & Fukuda, M. (1989) Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at *Alu* repeat sequences. *Proc. Natl Acad. Sci. USA*, 86, 4619–4623

Kushiro, J.-I., Hirai, Y., Kusunoki, Y., Kyoizumi, S., Kodama, Y., Wakisaka, A., Jeffreys, A., Cologne, J.B., Dohi, K., Nakamura, N. & Akiyama, M. (1992) Development of a flow-cytometric HLA-A locus mutation assay for human peripheral blood lymphocytes. *Mut. Res.*, 272, 17–29

Kyoizumi, S., Nakamura, N., Hakoda, M., Awa, A.A., Bean, M.A., Jensen, R.H. & Akiyama, M. (1989) Detection of somatic mutations at the glycophorin A locus in erythrocytes of atomic bomb survivors using a single beam flow sorter. *Cancer Res.*, 49, 581–588

Kyoizumi, S., Akiyama, M., Hirai, Y., Kusunoki, Y., Tanabe, K. & Umeki, S. (1990) Spontaneous loss and alteration of antigen receptor expression in mature CD4+ T cells. *J. Exp. Med.* 171: 1981–1999

Kyoizumi, S., Umeki, S., Akiyama, M., Hirai, Y., Kusunoki, Y., Nakamura, N., Endoh, K., Konishi, J., Sasaki, M.S., Mori, T., Fujita, S. & Cologne, J.B. (1992) Frequency of mutant T lymphocytes defective in the expression of the T-cell antigen receptor gene among radiation-exposed people. *Mut. Res.*, 265, 173–180

Langlois, R.G., Bigbee, W.L., Kyoizumi, S., Nakamura, N., Bean, M.A., Akiyama, M. & Jensen, R.H. (1987) Evidence for increased somatic cell mutations at the glycophorin A locus in atomic bomb survivors. *Science*, 236, 445–448

Langlois, R.G., Nisbet, B., Bigbee, W.L., Ridinger, D.N. & Jensen, R.H. (1990) An improved flow cytometric assay for somatic mutations at the glycophorin A locus in humans. *Cytometry*, 11, 513–521

Langlois, R.G., Akiyama, M., Kusunoki, Y., Dupont, B.R., Moore D.H., I.I., Bigbee, W.L., Grant, S.G. & Jensen, R.H. (1993) Analysis of somatic cell mutations at the glycophorin A locus in atomic bomb survivors: a comparative study of assay methods. *Rad. Res.*, 136, 11–117

Lipkowitz, S., Garry, V.F. & Kirsch, I.R. (1992) Interlocus V-J recombination measures genomic instability in agriculture workers at risk for lymphoid malignancies. *Proc. Natl Acad. Sci. USA*, 89, 5301–5305

Liu, Y., Hernandez, A.M., Shibata, D. & Cortopassi, G.A. (1994) *BCL2* translocation frequency rises with age in humans. *Proc. Natl Acad. Sci. USA*, 91, 8910–8914

Lubin, J.H., Boice, J.D., Jr, Edling, C., Hornung, R.W., Howe, G., Kunz, E., Kusiak, R.A., Morrison, H.I., Radford, E.P., Samet, J.M., Tirmarche, M., Woodward, A., Xiang, Y.S. & Pierce, D.A. (1994) *Radon and Lung Cancer Risk: A Joint Analysis of 11 Underground Miners Studies* (HIH Publication No. 94–3644), Bethesda, MD, US Department of Health and Human Services, Public Health Service, National Institutes of Health, pp 1–136

McCarron, M.A., Kutlaca, A. & Morley, A.A. (1989) The HLA-A mutation assay: improved technique and normal results. *Mut. Res.*, 225, 189–193

Manchester, D.K., Nicklas, J.A., O.'Neill, J.P., Lippert, M.J., Grant, S.G., Langlois, R.G., Moore, I.I., D.H., Jensen, R.H., Albertini, R.J. & Bigbee, W.L. (1995) Sensitivity of somatic mutations in human umbilical cord blood to maternal environments. *Environ. Mol. Mutagen.*, 26, 203–212

Mendelsohn, M.L. (1990) New approaches for biological monitoring of radiation markers. *Health Phys.*, 59, 23–28

Messing, K. & Bradley, W.E.C. (1985) *In vivo* mutant frequency rises among breast cancer patients after exposure to high doses of gamma radiation. *Mut. Res.*, 152, 107–112

Messing, K., Seifert, A.M. & Bradley, W.E.C. (1986) *In vivo* mutant frequency of technicians professionally exposed to ionizing radiation. In: Sorsa, M. & Norppa, H., eds, *Monitoring of Occupational Genotoxicants*, New York, Alan R. Liss, pp. 87–97

Messing, K., Ferraris, J., Bradley, W.E.C., Swartz, J. & Seifert, A.M. (1989) Mutant frequency of radiotherapy technicians appears to be associated with recent dose of ionizing radiation. *Health Phys.*, 57, 537–544

Morley, A.A., Trainor, K.J., Dempsey, J.L. & Seshardi, R.S. (1985) Methods for study of mutations and mutagenesis in human lymphocytes. *Mut. Res.*, 147, 363–367

Morley, A.A., Grist, S.A., Turner, D.R., Kutlaca, A. &

Bennett, G. (1990) Molecular nature of *in vivo* mutations in human cells at the autosomal HLA-A locus. *Cancer Res.*, 50, 4584–4587

Mott, M.G., Boyse, J., Hewitt, M. & Radford, M. (1994) Do mutations at the glycophorin A locus in patients treated for childhood Hodgkin's disease predict secondary leukaemia? *Lancet*, 343, 828–830

Nicklas, J.A., Hunter, T.C., O.'Neill, J.P. & Albertini, R.J. (1989) Molecular analyses of *in vivo HPRT* mutations in human T-lymphocytes: III. Longitudinal study of *HPRT* gene structural alterations and T-cell clonal origins. *Mut. Res.*, 215, 147–160

Nicklas, J.A., Falta, M.T., Hunter, T.C., O.'Neill, J.P., Jacobson-Kram, D., Williams, J. & Albertini, R.J. (1990) Molecular analysis of *in vivo HPRT* mutations in human lymphocytes V. Effects of total body irradiation secondary to radioimmunoglobulin therapy (RIT). *Mutagenesis*, 5, 461–468

Nicklas, J.A., O.'Neill, J.P., Hunter, T.C., Falta, M.T., Lippert, M.J., Jacobson-Kram, D., Williams, J.R. & Albertini, R.J. (1991) *In vivo* ionizing irradiations produce deletions in the *HPRT* gene of human T-lymphocytes. *Mut. Res.*, 250, 383–396

O'Neill, J.P., McGinniss, M.J., Berman, J.K., Sullivan, L.M., Nicklas, J.A. & Albertini, R.J. (1987) Refinement of a T-lymphocyte cloning assay to quantify the *in vivo* thioguanine-resistant mutant frequency in humans. *Mutagenesis*, 2, 87–94

Ostrosky-Wegman, P., Montero, R.M., Cortinas de Nova, C., Tice, R.R. & Albertini, R.J. (1988) The use of bromodeoxyuridine labeling in the human lymphocyte HGPRT somatic mutation assay. *Mut. Res.*, 191, 211–214

Ostrosky-Wegman, P., Montero, R., Palao, A., Cortinas de Nava, C., Hurtado, E. & Albertini, R.J. (1990) 6-thioguanine-resistant T-lymphocyte autoradiographic assay. Determination of variation frequencies in individuals suspected of radiation exposure. *Mut. Res.*, 232, 49–56

Ostrosky-Wegman, P., Gonsebatt, M.E., Montero, R., Vega, L., Barba, H., Espinosa, J., Palao, A., Cortinas, C., Garcia-Vargas, G., Del Razo, L.M. & Cebrian, M. (1991) Lymphocyte proliferation kinetics and genotoxic findings in a pilot study on individuals chronically exposed to arsenic in Mexico. *Mut. Res.*, 250, 477–482

Palmer, R.G., Smith-Burchnell, C.A., Pelton, B.K., Hylton, W. & Denman, A.M. (1988) Use of T cell cloning to detect *in vivo* mutations induced by cyclophoshamide. *Arth. Rheum.*, 31, 757–761

Partanen, R., Hemminki, K., Koskinen, H., Luo, J.-C., Carney, W.P. & Brandt-Rauf, P.W. (1994) The detection of increased amounts of the extracellular domain of the epidermal growth factor receptor in serum during carcinogenesis in asbestosis patients. *JOM*, 36, 1324–1328

Patel, P.I., Nussbaum, R.I., Framson, P.E., Ledbetter, D.H., Caskey, C.T. & Chinault, A.C. (1984) Organization of the *HPRT* gene and related sequences in the human genome. *Som. Cell Mol. Gen.*, 10, 483–493

Perera, F.P., Motzer, R.J., Tang, D., Reed, E., Parker, R., Warburton, D., O.'Neill, J.P., Albertini, R., Bigbee, W.L., Jensen, R.H., Santella, R., Tsai, W.-Y., Simon-Cereijido, G., Randall, C. & Bosl, G. (1992a) Multiple biologic markers in germ cell tumor patients treated with platinum-based chemotherapy. *Cancer Res.*, 52, 3558–3565

Perera, F.P., Hemminki, K., Gryzbowska, E., Motykiewicz, G., Michalska, H., Satella, R.M., Young, T.L., Dickey, C., Brandt-Rauf, P.W., DeVivo, I., Blaner, W., Tsai, W.-Y. & Chorazy, M. (1992b) Molecular and genetic damage in humans from environmental pollution in Poland. *Nature*, 35, 256–258

Perera, F.P., Tang, D.L., O.'Neill, J.P., Bigbee, W.L., Albertini, R.J., Santella, R., Ottman, R., Tsai, W.-Y., Dickey, C., Mooney, L.A., Savela, K. & Hemminki, K. (1993) HPRT and glycophorin A mutations in foundry workers: relationship to PAH exposure and to PAH-DNA adducts. *Carcinogenesis*, **14**, 969–973

Perera, F.P., Dickey, C., Santella, R., O.'Neill, J.P., Albertini, R.J., Ottman, R., Tsai, W.Y., Mooney, L.A., Savela, K. & Hemminki, K. (1994) Carcinogen-DNA adducts and gene mutation in foundry workers with low-level exposure to polycyclic aromatic hydrocarbons. *Carcinogenesis*, 15, 2905–2910

Pluth, J.M., Nicklas, J.A., O.'Neill, J.P. & Albertini, R.J. (1996) Increased frequency of specific genomic deletions resulting from *in vitro* malathion exposure. *Cancer Res.*, 56, 2393–2399

Rainville, I.R., Albertini, R.J., O.'Neill, J.P. & Nicklas, J.A. (1995) Breakpoints and junctional regions of intragenic deletions in the *HPRT* gene in human T-cells. *Som. Cell Mol. Gen.*, 21, 309–326

Robinson, D., Goodall, K., Albertini, R.J., O.'Neill, J.P., Finette, B., Sala-Trepat, M., Moustacchi, E., Tates, A.D., Beare, D.M., Green, M.H.L. & Cole, J. (1994) An analysis of *in vivo HPRT* mutant frequency in circulating T-lymphocytes in the normal human population: a comparison of four databases. *Mut. Res.*, 313, 227–247

Rothman, N., Haas, R., Hayes, R.B., Li, G.-L., Wiemels, J., Campleman, S., Quintana, P.J.E., Xi, L.-J., Dosemeci, M., Titenko-Holland, N., Meyer, K.B., Lu, W., Zhang, L.P., Bechtold, W., Wang, Y.-Z., Kolachana, P., Yin, S.-N., Blot, W. & Smith, M.T. (1995) Benzene induces gene-duplicating but not gene-inactivating mutations at the glycophorin A locus in exposed humans. *Proc. Natl Acad. Sci. USA*, 92, 4069–4073

Sala-Trepat, M., Cole, J., Green, M.H.L., Rigaud, O., Vilcoq, J.R. & Moustacchi, E. (1990) Genotoxic effects of radiotherapy and chemotherapy on the circulating lymphocytes of breast cancer. III: Measurement of mutant frequency to 6-thioguanine resistance. *Mutagenesis*, 5, 593–598

Sala-Trepat, M., Boyse, J., Richard, P., Papadopoulo, D. & Moustacchi, E. (1993) Frequencies of *HPRT*- lymphocytes and glycophorin A variants erythrocytes in Fanconi anemia patients, their parents and control donors. *Mut. Res.*, 289, 115–126

Seifert, A.M., Bradley, W.C. & Messing, K. (1987) Exposure of nuclear medicine patients to ionizing radiation is associated with rises in HPRT- mutant frequency in peripheral T-lymphocytes. *Mut. Res.*, 191, 57–63

Seifert, A.M., Demers, C., Debeau, H. & Messing, K. (1993) HPRT-mutant frequency and lymphocyte characteristics of workers exposed to ionizing radiation on a sporadic basis: a comparison of two exposure indicators, job title and dose. *Mut. Res.*, 319, 61–70

Stamatoyannopoulos, G.P. & Nute, P.E. (1981) Screening of human erythrocytes for products of somatic mutation, an approach and critique. In: Hook, E.B. & Porter, I.H., eds, *Population and Biological Aspects of Human Mutation*, New York, Academic Press, pp. 265–273

Stamatoyannopoulos, G., Nute, P., Lindsley, D., Farguhar, M.B., Nakamato, B. & Papayannepoulou, T.N. (1984) Somatic-cell mutation monitoring system based on human hemoglobin mutants. In: Ansari, A.A. & de Serres, F.J., eds, *Single Cell Monitoring System, Topics in Chemical Mutagenesis* (Vol 3), New York, Plenum Press, pp. 1–35

Stark, M.H., Tucker, J.H., Thomson, E.J. & Perry, P.E. (1984) An automated image analysis system for the detection of rare autoradiographically labeled cells in the human lymphocyte HGPRT assay. *Cytometry*, 5, 250–257

Stout, J.T. & Caskey, C.T. (1985) Hprt: gene structure, expression, and mutation. *Ann. Rev. Genet.*, 19, 127–148

Straume, T., Langlois, R.G., Lucas, J., Jemsem, R.H., Bigbee, W.L., Ramalho, A.T. & Brandao-Mello, C.E. (1991) Novel biodosimetry methods applied to victims of the Goiania accident. *Health Phys.*, 60, 71–76

Strauss, G.H. & Albertini, R.J. (1977) 6-thioguanine resistant lymphocytes in human perpheral blood. In: Scott, D., Bridges, B.A. & Sobels, F.H., eds, *Progress in Genetic Toxicology: Developments in Toxicology and Environmental Sciences* (vol. 2), Amsterdam, Elsevier/North Holland Biomedical Press, p. 327

Strauss, G.H. & Albertini, R.J. (1979) Enumeration of 6-thioguanine resistant peripheral blood lymphocytes in man as a potential test for somatic cell mutations arising *in vivo*. *Mut. Res.*, 61, 353–379

Tates, A.D., Bernini, L.F., Natajaran, A.T., Ploem, J.S., Verwoed, N.P., Cole, J., Green, M.H.L., Arlett, C.F. & Norris, P.N. (1989) Detection of somatic mutations in man: HPRT mutations in lymphocytes and hemoglobin mutations in erythrocytes. *Mut. Res.*, 213, 73–82

Tates, A.D., Grummt, T., Tornqvist, M., Farmer, P.B., van Dam, F.J., van Mossel, H., Schoemaker, H.M., Osterman-Golkar, S., Uebel, C., Tang, Y.S., Zwinderman, A.H., Natarajan, A.T. & Ehrenberg, L. (1991) Biological and chemical monitoring of occupational exposure to ethylene oxide. *Mut. Res.*, 250, 483–497

Tates, A.D., van Dam, F.J., Natarajan, A.T., Zwinderman, A.H. & Osanto, S. (1994a) Frequencies of HPRT mutants and micronuclei in lymphocytes of cancer patients under chemotherapy: a prospective study. *Mut. Res.*, 307, 293–306

Tates, A.D., Grummit, T., van Dam, F.J., de Zwart, F., Kasper, F.J., Rothe, R., Stirn, H., Zwinderman, A.H. & Natarajan, A.T. (1994b) Measurement of frequencies of HPRT mutants, chromosomal aberrations, micronuclei, sister-chromatid exchanges and cells with high frequencies of SCEs in styrene/dichloromethane-exposed workers. *Mut. Res.*, 313, 249–262

Tates, A.D., Boogaard, P.J., Darroudi, F., Natarajan, A.T., Caubo, M.E. & van Sittert, N.J. (1995) Biological effect monitoring in industrial workers following incidental exposure to high concentrations of ethylene oxide. *Mut. Res.*, 329, 63–77

Tates, A.D., van Dam, F.J., de Zwart, F.A., Darroudi, F., Natarajan, A.T., Rossner, P., Peterkova, K., Peltonen, K., Demopoulos, N.A., Vlachodimitropoulos, D. & Sram, R.J. (1996) Biological effect monitoring in industrial workers from the Czech Republic that were exposed to low levels of butadiene. *Toxicology*, 113, 91–99

Theocharis, S., Sfikakis, P.P., Lipnick, R.N., Klipple, G.L., Steinberg, A.D. & Tsokos, G.C. (1995) Characterization of *in vivo* mutated T cell clones from patients with systemic lupus erythematosus. *Clin. Immunol. Immunopath.*, 74, 135–142

Turner, D.R., Grist, S.A., Janatipour, M. & Morley, A.A. (1988) Mutations in human lymphocytes commonly involve gene duplication and resemble those seen in cancer cells. *Proc. Natl Acad. Sci. USA*, 85, 3189–3193

Tycko, B. & Sklar, J. (1990) Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells*, 2, 1–8

Umeki, S., Kyoizumi, S., Kusunoki, Y., Nakamura, N., Sasaki, M., Mori, T., Ishikawa, Y., Cologne, J.B. & Akiyama, M. (1991) Flow cytometric measurements of somatic cell mutations in Thorotrast patients. *Jpn J. Cancer Res.*, 82, 1349–1353

van Dam, F.J., Camps, J.A., Woldring, V.M., Natarajan, A.T., van der Wall, E.E., Zwinderman, A.H., Lohman, P.H.M., Pauwels, E.K.J. & Tates, A.D. (1991) Radionuclide

angiography with Technetium-99m *in vivo* labeled erythrocytes does not lead to induction of miutations in the HPRT gene of human T-lymphocytes. *J. Nucl. Med.*, 32, 814–818

Vijayalaxmi, Wunder, E. & Schroeder, T.M. (1985) Spontaneous 6-thioguanine-resistant lymphocytes in Fanconi anemia patients and their heterozygous parents. *Hum. Genet.*, 70, 264–270

Vodicka, P., Bastlova, T., Vodickova, L., Peterkova, K., Lambert, B. & Hemminki, K. (1995) Biomarkers of styrene exposure in lamination workers: levels of 06-guanine DNA adducts, DNA strand breaks and mutant frequencies in the hypoxanthine guanine phosphoribosyltransferase gene in T-lymphocytes. *Carcinogenesis*, 16, 1473–1481

Ward, J.B., Ammenheuser, M.M., Bechtold, W.E., Whorton, E.B., Jr & Legator, M.S. (1994) *HPRT* mutant lymphocyte frequencies in workers at a 1,3-butadiene production plant. *Environ. Health Perspect.* 102, 79–85

Ward, J.B., Jr, Ammenheuser, M.M., Whorton, E.B., Jr, Bechtold, W.E., Kelsey, K.T. & Legator, M.S. (1996) Biological monitoring for mutagenic effects of occupational exposure to butadiene. *Toxicology*, 113, 84–90

Corresponding author

R.J. Albertini

Genetics Laboratory, University of Vermont, 32 N. Prospect St, Burlington, VT 05401-0508, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Cytogenetic end-points as biological dosimeters and predictors of risk in epidemiological studies

J.D. Tucker, D.A. Eastmond and L.G. Littlefield

Cytogenetic end-points have been successfully used in epidemiological studies for many years. Conventional end-points are now being replaced by procedures that utilize molecular methods, with greatly increased sensitivity, specificity and precision. In this paper we briefly review the most common cytogenetic assays that are useful in epidemiological settings, including structural chromosome aberrations, micronuclei, sister chromatid exchanges and analysis of interphase cells for aneuploidy. We describe new developments of each assay, where applicable, and discuss the strengths and weaknesses of the assays for detecting exposures and estimating risks. Finally, pertinent information concerning each of the assays that is useful in designing epidemiological studies is summarized in a table. It is hoped that the information presented here will be useful to individuals who are interested in applying biomarkers to studies of human environmental exposure and disease.

Several cytogenetic end-points have potential use as biomarkers of exposure to clastogens. These include structural chromosome aberrations, micronuclei and sister chromatid exchanges (SCEs). Structural aberrations have been studied for more than half a century and thousands of papers have been published on the effects of DNA-damaging agents in inducing chromosome damage. At the molecular level, aberrations are believed to result from double-strand breaks in DNA that remain unrepaired or that undergo aberrant rejoining, giving rise to chromosomal rearrangements.

While dozens of chemical, physical and biological agents induce DNA damage, ionizing radiation is among the most efficient in inducing double-strand breaks that lead to the prompt formation of chromosome aberrations. It has been known for more than three decades that radiation induces dose-dependent increases in asymmetrical (i.e. unstable) and symmetrical (i.e. stable) chromosome aberrations in cells throughout the body, including mature cells and progenitor stem cells in the haematopoietic lineage. Induced aberrations in human lymphocytes have been used as biological dosimeters to gauge exposure levels since the early 1960s. Information gleaned from a multitude of radiation cytogenetic studies has also provided an insight into the mechanisms that govern the induction, persistence, accumulation and elimination of chromosome damage and into the relevance of such damage to risk of late effects in exposed populations. Attempts to derive recommendations regarding the applicability of cytogenetic techniques in population epidemiology must take into account a wealth of information that has been accumulated over the past 30 years. One of the compelling arguments for studying cytogenetics is the relationship between chromosome aberrations and tumour formation. It is of particular interest to discover whether aberration frequencies can be used to predict cancer. Several recent reports (Hagmar *et al.*, 1994; Bonassi *et al.*, 1995) have shown that aberration frequencies are increased prior to the clinical manifestation of disease.

If this relationship holds true upon further investigation, then it will be important to determine the statistical strength of this predictability for various types of cancer, and to decipher the biological mechanisms responsible for these intriguing observations.

Micronuclei contain chromosomes or portions of chromosomes that exist separately from the main nucleus of a cell. Their existence has been known for decades, and they are frequently used to quantify exposure to chemicals and radiation. Micronuclei can be observed in almost any cell type, and for this reason many variations of the assay exist. Compared to other cytogenetic assays, there are several advantages of quantifying micronuclei, including speed and ease of the analysis, and especially the absence of the requirement for examining cells in metaphase. Thus, many cell types are amenable to analysis of micronuclei, and most can be employed in epidemiological studies. Several mechanisms of action contribute to the formation of micronuclei (Heddle *et al.*, 1983), including chromosome breakage (clastogenesis) and spindle disruption (aneuploidogenesis).

SCEs are the cytological manifestation of interchanges between DNA replication products at apparently homologous loci. SCEs have been commonly employed to evaluate cytogenetic responses to chemical exposure, and hundreds of chemicals have been evaluated in a wide variety of in-vitro and in-vivo short-term experiments (Tucker *et al.*, 1993a). Numerous studies have also been conducted to evaluate possible environmental exposure in humans. Unfortunately, the mechanism of SCE formation remains unknown, thus limiting the use of this biomarker for evaluations concerning the effects of exposures.

The objective of this article is to provide pertinent information concerning these cytogenetic assays that is applicable when designing epidemiological studies of human environmental exposure and disease. Table 1 summarizes the salient features of these end-points with respect to their utility as biomarkers in such studies.

Structural chromosome aberrations

Since the early 1960s, three methods have been used to observe and quantify structural chromosome aberrations in human cells. These are analysis of unbanded chromosomes, banded chromosomes and painting. Classical studies typically used DNA dyes such as Giemsa to stain all chromosomes in the complement, and the earliest methods did not allow unique identification of all the chromosomes in the complement. Unstable aberrations such as dicentrics and rings are readily detected using these techniques. Symmetrical chromosome rearrangements are not easily observed, and early studies used clinical laboratory methods of idiogramming or karyotyping of photographic prints to quantify translocation frequencies (e.g. Buckton *et al.*, 1962; Littlefield & Joiner, 1978). Each metaphase cell had to be counted and mentally karyotyped by locating and comparing the sizes and centromere location of chromosome pairs 1, 2, 3 and 16, and by counting and comparing the centromere locations of the B, D, F and G group chromosomes. Typically, any metaphase cell that contained an abnormal monocentric chromosome was photographed and karyotyped. Interstitial translocations and those involving exchanges of small pieces of chromatin were not detected. Comparisons of translocations detected by these methods and even more laborious banding techniques have demonstrated that, when performed by highly skilled staff, group analysis techniques detect upwards of 75% of the cells with symmetrical rearrangements in the entire genome (Buckton, 1976; Ohtaki *et al.*, 1982). Group analysis and banding methods are slow and tedious and require highly-trained staff who have complete familiarity with the human chromosome complement. Thus, the methodology is not applicable for the large-scale screening of populations that is frequently of interest in epidemiological surveillance.

The application of cytogenetic techniques in population epidemiology is becoming more feasible with the recent development of fluorescence in-situ hybridization (FISH) using chromosome-specific DNA probes that paint specific chromosome pairs along their entire lengths (e.g. Lucas *et al.*, 1989; Natarajan *et al.*, 1991; Tucker *et al.*, 1993b). These methods allow the selective identification of each pair of chromosomes in the human genome and permit rapid recognition of chromosome breakage and exchange events between painted and non-painted chromosomes. Symmetrical exchanges can be readily observed, and these new methods have refocused the attention of the scientific community on the translocation as a potential retrospective biomarker of exposure. Because chromosome painting techniques typically quantify only those exchanges that occur between painted and non-painted chromosomes,

it must be assumed that the observed exchange events in the target chromosomes are accurate surrogate measures of those occurring in the entire genome. Estimates of total genomic frequencies can be calculated (Lucas *et al.*, 1992) by taking into consideration the percentage of the total chromosomal DNA that is painted, and by assuming that random breakage and exchange occurred between all pairs of chromosomes. In-situ hybridization procedures require some knowledge of molecular biology techniques, expensive reagents and a good-quality fluorescence microscope. Procedures for hybridization typically result in some loss of chromosomal morphology, as a result of subjecting cells to extremes in temperatures, high concentrations of salts, as well as enzymes and formamide, which can cause problems in ascertaining accurately some of the aberration types. None the less, painting methods offer a number of advantages in terms of population monitoring, including the relative ease of scoring large numbers of metaphase cells to obtain estimates of translocation frequencies, objectivity of scoring (which is expected to reduce interscorer variability), and the potential for full computerization of data collection, which would make large-scale monitoring a feasibility.

Prior to the publication of painting methods, dozens of cytogenetic studies were conducted using classical metaphase analysis to evaluate the induction and persistence of structural chromosome aberrations in various types of somatic cells in individuals and populations with medical, occupational or accidental exposures. A number of conclusions derived from these studies have now been verified and expanded upon with chromosome painting techniques. In the following paragraphs we summarize the current status of knowledge regarding the applications and limitations of structural chromosomal aberrations as prompt or retrospective biomarkers of exposure and predictors of risk in epidemiology studies.

Sensitivity of structural aberrations for detecting recent exposures

Studies of radiation-induced chromosome aberrations in hundreds of persons with accidental radiation exposures have firmly established that unstable (or asymmetrical) types of chromosome aberrations in cultured peripheral blood lymphocytes are the most sensitive biological end-points currently available for detecting recent exposures to clastogenic agents (e.g. Bender *et al.*, 1988). Unstable aberrations such as dicentrics are easily identified in metaphase cells of cultured peripheral blood lymphocytes, and because these occur with low background frequency (~1–2 per thousand) in lymphocytes of persons with low levels of environmental or occupational exposures to clastogens, relatively minor increases above background can be readily detected. For example, when sufficient numbers of metaphase cells are evaluated from lymphocyte cultures initiated within a few weeks of exposure, scoring of dicentrics can readily detect recent whole-body exposures of ~10–20 cGy in individuals and ~5 cGy in populations exposed to ionizing radiation. Dicentrics and other types of unstable aberrations are lethal when cells undergo division, and are therefore quickly eliminated from proliferating cell compartments such as bone marrow. Dicentrics will only be observed in populations of mature lymphocytes that have not undergone in-vivo cell divisions in the interim between exposure and collection of blood for cytogenetic evaluation. The average life expectancy of T-lymphocytes appears to be bimodal, with means of approximately 1.1 and 6.3 years (Bogen, 1993), and lymphocytes carrying dicentrics will be useful for biological dosimetry purposes. Thus, in terms of population monitoring, unstable aberrations in lymphocytes are useful as transient biomarkers of recent exposure and reflect only a snapshot of a person's exposure history. In general, samples obtained within 6 weeks of a single acute exposure should be suitable for biodosimetric analysis by scoring dicentric chromosomes. After this time, dicentric frequencies begin to decline and precise dosimetry becomes more difficult. Back-extrapolation of dicentric frequencies obtained after 6 weeks may be possible, but additional sources of error are introduced. Measurements of translocation frequencies become the preferred method for quantifying exposure due to the stability of the lesions.

Translocations as retrospective biodosimeters

As first observed and discussed in the early 1960s, when persons are exposed to physical or chemical agents capable of inducing chromosome damage, stable chromosome rearrangements, such as reciprocal translocations, are also induced in mature cells and in progenitor cells in various stem cell

Table 1. Summary of considerations for sample acquisition in epidemiological studies

	Bioassay					
	Structural chromosome aberrations					
Tissue	Unbanded	Banded	Painting	Sister chromatid exchanges	Micronuclei	Interphase (aneuploidy)
Peripheral blood						
Assay	Routine	Routine	Routine	Routine	Nucleated cells: routine Erythrocytes: splenectomy required	Nucleated cells: routine Erythrocytes: not yet done, splenectomy would be required
Sample acquisition	Phlebotomy	Phlebotomy	Phlebotomy	Phlebotomy	Phlebotomy (finger-prick for erythrocytes)	Phlebotomy
Sample storage and staining considerations	No special requirements	Slides should be banded quickly	Store at –20°C in N_2 atmosphere	No special requirements	Store at –20°C in N_2 atmosphere	Store at –20 °C in N_2 atmosphere
Bone marrow						
Assay	Possible	Possible	Possible	Not practical[a]	Possible	Possible
Sample acquisition	Difficult, painful, invasive	Difficult, painful, invasive	Difficult, painful, invasive		Difficult, painful, invasive	Difficult, painful, invasive
Sample storage and staining considerations	As for blood	As for blood	As for blood		As for blood	As for blood
Fibroblasts						
Assay	Possible	Possible	Possible	Possible	Possible	Possible
Sample acquisition	Leaves small scar	Leaves small scar	Leaves small scar	Leaves small scar	Leaves small scar	Leaves small scar
Sample storage and staining considerations	As for blood	As for blood	As for blood	As for blood	As for blood	As for blood

Table 1. (Contd) Summary of considerations for sample acquisition in epidemiological studies

	Bioassay					
	Structural chromosome aberrations			Sister chromatid exchanges	Micronuclei	Interphase (aneuploidy)
Tissue	**Unbanded**	**Banded**	**Painting**			
Buccal/nasal mucosa						
Assay	Not practical[b]	Not practical[b]	Not practical[b]	Not practical[b]	Possible	Possible
Sample acquisition					Rinse mouth, scrape off cells with tongue depressor or toothbrush	Rinse mouth, scrape off cells with tongue depressor or toothbrush
Sample storage and staining considerations					As for blood	As for blood
Urinary tract epithelium						
Assay	Not practical[b]	Not practical[b]	Not practical[b]	Not practical[b]	Possible	Possible
Sample acquisition					Urine sample[c]	Urine sample[c]
Sample storage and staining considerations					Tough membranes require treatment with proteases to facilitate probe penetration	Tough membranes require treatment with proteases to facilitate probe penetration
Sperm						
Assay	Possible[d]	Possible[d]	Possible[d]	Not possible[a]	Possible	Routine
Sample acquisition	Ejaculate	Ejaculate	Ejaculate		Ejaculate	Ejaculate
Sample storage and staining considerations	No special requirements	Slides should be banded quickly	Store at –20°C in N_2 atmosphere		Store as for blood[e]	Store as for blood[e]

[a]Bromodeoxyuridine (BrdU) is required in the penultimate S-phase of the cell cycle for SCEs to be visualized. Accomplishing this requires culturing the cells or in-vivo administration of BrdU, in which case the analysis of other tissues, such as peripheral blood, becomes more feasible.

[b]These assays require cells in metaphase, which is difficult to accomplish for these cell types.

[c]First morning void should not be collected. Collect multiple voids from males and one or more from females. Cell yield is much higher in females, but cells originate from both the urinary and reproductive tracts. In males, they are primarily from the urinary tract. Many cells from males are apoptotic or have poor nuclear morphology.

[d]Analysis of sperm chromosomes in metaphase is possible following fusion with hamster eggs (e.g. Brandriff *et al.*, 1988).

[e]Wash or smear directly on slide. Nuclei are extremely compact and must be swelled to facilitate probe penetration. May also require treatment with multiple proteases (Robbins *et al.*, 1993).

compartments. Because these types of rearrangements typically do not impede cellular progression through mitosis, translocations are passed on to daughter cells during cell proliferation. If cells bearing translocations do not have a selective advantage or disadvantage, compared to those with normal chromosome complements, then one might expect them to persist throughout a lifetime and to reflect broadly the person's cumulative exposure to clastogens (Kleinerman *et al.*, 1989).

Much information on the behaviour of these persistent stable types of chromosome aberrations has been gleaned from long-term follow-up studies of hundreds of persons accidentally exposed to ionizing radiation, or exposed during various medical procedures. Evaluations using classical methods and painting techniques have demonstrated that the majority of stable aberrations observed in lymphocytes are reciprocal or non-reciprocal translocations (e.g. Ishihara & Kumatori, 1969; Awa, 1974; Littlefield & Joiner, 1978; Tucker *et al.*, 1994; Ramsey *et al.*, 1995). In studies of relatively large cohorts of non-exposed persons, translocation frequencies observed in cultured lymphocytes are several times higher and considerably more variable than dicentric frequencies. Data from classical analyses have shown that mean frequencies of stable aberrations in lymphocytes are higher in groups of older women (Littlefield *et al.*, 1991), and recent data from painting studies have shown a strong age effect in controls (Tucker *et al.*, 1994; Ramsey *et al.*, 1995). The fact that variable background frequencies are observed in normal controls is a confounder that must be taken into consideration when attempting to gauge the sensitivity of translocation analyses as an index of exposure.

For irradiated populations, increased frequencies of lymphocytes with stable aberrations are observed for upwards of 40 or more years after exposure, and for any individual, the numbers of lymphocytes bearing radiation-induced stable aberrations remain relatively constant over time (e.g. Buckton *et al.*, 1962; Awa, 1974). Infrequently, three or more metaphase cells with apparently identical translocations or insertions have been observed in individuals in several study cohorts, including A-bomb survivors, ankylosing spondylytis patients, Y-12 accident survivors, persons exposed at Chernobyl (Salassidis *et al.*, 1995), as well as in controls (Tucker *et al.*, 1994). These observations suggest that progenitor subsets of T-lymphocytes may undergo selective clonal expansion under certain conditions. Studies in populations with exposures over a wide range of doses (i.e. A-bomb survivors in Hiroshima and Nagasaki, and women irradiated locally for cervical cancer) have demonstrated that the mean frequencies of lymphocytes bearing translocations show excellent regression with radiation dose for upwards of four decades after exposure (Awa, 1974; Kleinerman *et al.*, 1989). Thus, stable aberrations appear to be sensitive retrospective biomarkers of exposure levels for populations as a whole. However, because considerable variability has been observed between persons within dose groups, there are questions regarding the precision of stable aberrations in lymphocytes as retrospective dosimeters for individuals studied many years after exposure (Kleinerman *et al.*, 1989). As discussed, clonal expansion of subsets of memory T-cells in response to random antigenic challenge would tend to increase interperson variability in the observed frequencies of cells with radiation-induced translocations within dose groups with increasing time after a specific exposure.

Translocations as predictors of risk of late effects

As early as 1962, consideration was given to the possible relationship between persistent chromosome rearrangements in lymphocytes and the risk of late effects (Buckton *et al.*, 1962). Recently, it has been postulated that the burden of stable chromosomal rearrangements in lymphocytes may be useful as a surrogate biomarker that gauges the relative levels of cumulative spontaneous or agent-induced genetic damage sustained by the haematopoietic stem cell compartment (Littlefield *et al.*, 1991). Because chromosomal rearrangements are thought to be relevant to, and play a role in, carcinogenesis (e.g. Heim & Mitelman, 1987; Sandberg, 1990; Cleary, 1991; Stanbridge, 1992), it can be reasoned that the frequencies of stable aberrations in somatic cells would be positively correlated with the risk for cancer in exposed populations.

A major prospective study by the Nordic Study Group is underway, the purpose of which is to gather information on the correlation between various cytogenetic end-points and future cancer risk (Brogger *et al.*, 1990; Hagmar *et al.*, 1994). Cytogenetic studies in this cohort were conducted

between 1970 and 1988 and aberrations include chromatid as well as chromosome breaks and exchanges. Eighty-five cancers were diagnosed in the Nordic Study Group cohort during the period 1970–1991. In the second follow-up report (Hagmar *et al.*, 1994), no significant trend in the standardized incidence ratio was observed for either sister chromatid exchanges or micronuclei, but there was a statistically significant linear trend for chromosomal aberrations with regard to cancer risk. Similar results have been seen by Bonassi *et al.* (1995).

In another study, attempts were made to derive correlations between stable radiation-induced chromosome aberrations in lymphocytes and cancer risk in women who received high-dose, localized radiotherapy for cervical cancer or low-dose exposures to treat benign gynaecological disease (Kleinerman *et al.*, 1994). Leukaemia risk is nearly the same in both groups, despite a 10-fold difference in average dose to their bone marrow. The cervical cancer patients received fractionated, high-dose-rate radiotherapy, with average accumulated doses of up to 40 Gy to the whole pelvis, and it has been postulated that their lower excess leukaemia risk per unit dose may be attributed to cell killing in the high-dose fields, with wasted radiation effect. Classical studies using group analyses conducted nearly two decades after exposure demonstrated that the rate of stable translocations in lymphocytes was only slightly higher among the cervical cancer patients. Assuming that the fraction of cytogenetically aberrant stem cells that survive radiotherapy contribute to the leukaemogenic process, then the cytogenetic findings are consistent with epidemiological findings of comparable overall leukaemia risks seen in these two irradiated populations. These data further suggest that in instances of high-dose localized exposures, persistent translocations in lymphocytes appear to serve as biomarkers of effective risk as well as biomarkers of dose.

Cytogenetic evaluations in each of these cohorts were undertaken prior to the introduction of painting techniques, and thus a large data base has not yet been accumulated that would allow prospective studies to determine the correlation between the burden of translocations in somatic cells and the risk of late effects. It is likely that most translocations will damage or interrupt the sequence of at least one, if not two, genes. Assuming the average gene is ~20–25 kb long, including introns, and assuming 100 000 genes and 3.3×10^9 base pairs/haploid genome, it is obvious that no more than ~25–40% of the genome occupies the spaces between the genes. Although a small fraction of all genes appears to be important for tumorigenesis, DNA-damaging agents that increase the total burden of translocations in somatic cells would have a finite probability of inducing genetic alterations at sites that are relevant in carcinogenesis. Chromosome painting techniques will now allow for large-scale cytogenetic epidemiology studies of populations known to be at risk for cancer. Such studies are needed and are likely to produce important information regarding the association between induced chromosome aberrations in somatic cells and subsequent cancer risk.

Micronuclei

The existence of micronuclei has been known for many years, and they are frequently used to quantify exposure to chemicals and radiation. Micronuclei can be observed in almost any cell type, and for this reason many variations of the assay exist. Micronuclei can be evaluated in cell types that are not amenable to the analysis of structural aberrations, although the most common cells examined do belong to the haematopoietic system, including lymphocytes and erythrocytes. One caveat, however, is noteworthy. In humans, erythrocyte micronuclei persist only in splenectomized individuals because the spleen efficiently filters out erythrocytes containing micronuclei. Nevertheless, sufficient numbers of splenectomized people exist, at least in countries where surgical treatment of traumatic injuries is common, so that some limited studies involving these people have been performed (e.g. Schreunemachers & Everson, 1991). Other cell types that can be examined in humans are urothelial cells and exfoliated cells from the buccal and nasal mucosa. In humans, studies involving bone marrow and other internal tissues are technically possible, but because their acquisition requires invasive procedures they are not commonly employed. Peripheral lymphocytes are a commonly examined tissue, and may be evaluated for micronuclei with or without culturing. If the cells are cultured, cytochalasin B may be used to inhibit cytokinesis, and micronuclei are

scored in binucleated cells. If care is taken to evaluate only those cells with intact membranes, the investigator can be certain that all the products of a mitosis are present, which has important implications for studies where the mechanism of micronucleus formation is of interest.

Several mechanisms of action contribute to the formation of micronuclei (Heddle *et al.*, 1983), including chromosome breakage (clastogenesis) and spindle disruption (aneuploidogenesis). Until a few years ago, cytological evaluation of micronuclei provided little or no information concerning the type of damage that led to their formation. There was considerable debate concerning whether the size of a micronucleus could be used to elicit mechanisms. The basic premise was that larger micronuclei were caused by spindle disruption and would contain whole chromosomes, while smaller micronuclei would consist of one or more chromosome fragments. While the relationship between micronucleus size and its mechanism of origin may be generally true, we now know that there is a significant amount of overlap between these classes of micronuclei, with the result that it is impossible to know for certain how a single micronucleus arose.

The solution to this problem has been to use molecular methods to identify the contents of micronuclei. These methods consist of two general approaches: anti-kinetochore antibodies and DNA probes. Anti-kinetochore antibodies bind to the site of spindle fibre attachment to the chromosomes and thus serve as a marker for the centromere. Micronuclei with a kinetochore signal can reasonably be assumed to contain one or more whole chromosomes (the number of which can often be determined) and are indicative of processes that disrupt the mitotic spindle. Kinetochore-negative micronuclei are indicative of clastogenic processes (e.g. Eastmond & Tucker, 1989). A consensus sequence repetitive pan-centromeric DNA probe has been used to make similar determinations (e.g. Titenko-Holland *et al.*, 1994). Unique sequence DNA probes and repetitive probes for a single chromosome type can also be used to investigate the contents of micronuclei. For example, several studies have shown that the frequency of chromosomes in micronuclei is non-random. In particular, the X and Y chromosomes appear in micronuclei of females and males, respectively, far more often than expected by chance (Hando *et al.*, 1994; Nath *et al.*, 1995).

Micronuclei and cancer risk assessment.
The relationship between micronuclei and adverse health risks is not as well substantiated as it is for structural aberrations. The primary reason is that micronuclei are formed by more than one mechanism, and unless the contents of each micronucleus are known (especially with respect to the presence or absence of centromeres), the data may be subject to errors of interpretation. Furthermore, even when using micronuclei as a biomarker to evaluate exposure to an agent with a known mechanism, a method should be employed that enables discrimination between classes of micronuclei. There are several reasons for this. First, few agents have a single mechanism of action. Even known spindle disrupters appear to have some clastogenic activity, and radiation induces a significant number of kinetochore-positive micronuclei (Eastmond & Tucker, 1989). Thus, understanding the types and distribution of micronuclei will provide a more full description of the mechanism(s) of action of the compound in question. For agents with unknown or mixed mechanisms of action, evaluating the type(s) of micronuclei induced can lead to important insights into risks resulting from exposure. The second reason to use centromere detection methods is that spontaneous micronuclei arise by a variety of mechanisms and, as already mentioned, the contents of micronuclei are not a random representation of the genome. Understanding the types of micronuclei in the unexposed cells will enable a more refined statistical determination of exposure. For example, for analyses limited to clastogenic processes, only those micronuclei in the unexposed (control) group that lack a centromere need to be considered when determining whether the compound induced micronuclei containing chromosome fragments. This will have the effect of increasing the statistical power of the analysis.

To date, only a few studies have used molecular approaches to label centromeres of micronuclei in human populations (Hando *et al.*, 1994; Titenko-Holland *et al.*, 1994; Catalan *et al.*, 1995; Nath *et al.*, 1995) and none of the subjects had received an occupational or environmental exposure. The reasons may simply be that these methods require

additional effort and that not every cytogenetics laboratory is equipped for the molecular aspects of the work. In spite of these limitations, the advantages of including molecular methods in micronucleus analyses are significant and would substantially improve the ability of micronucleus assays to serve as a meaningful biomarker. As more micronucleus data including centromere information are gathered, so our understanding of the usefulness of the data for cancer epidemiology will improve. In the meantime, the analysis of micronucleus data can be performed in a manner which assumes that all micronuclei are formed either by clastogenic or by spindle disruptive processes. Although this will lead to an overestimation of the risk estimate, at least an upper confidence bound could be obtained. In spite of some limitations, micronuclei can be very useful as a biomarker.

Sister chromatid exchanges (SCEs)

SCEs can be measured in any eukaryotic cell that can be grown in bromodeoxyuridine (or a suitable alternative) for two cell cycles and examined in metaphase. SCEs have been evaluated in numerous studies involving human exposure. The assay yields quantifiable data from every metaphase cell, and the genotoxic potential of a chemical can be determined more rapidly than with some other cytogenetic methods (SCEs are not a good indicator of exposure to ionizing radiation). Besides the disadvantage of an unknown mechanism, the use of SCEs has been criticized because significant increases have been obtained from compounds such as NaCl and KCl which clearly are not expected to be genotoxic (Galloway *et al.*, 1987). Thus, a positive SCE response does not necessarily mean that a compound is genotoxic. However, SCE assays remain widely used for evaluating the ability of chemicals to induce genetic damage. Unlike some types of structural aberrations (especially reciprocal translocations), SCEs evaluated following in-vivo exposure show limited persistence and accumulation, which appears to reflect DNA repair processes as well as cell turnover.

Nevertheless, SCE data do have some appealing characteristics. The distribution of SCEs per cell is close enough to normal to allow parametric statistical analyses in most situations. Large SCE databases have been generated, making it possible to compare relative potencies for hundreds of compounds (Tucker *et al.*, 1993a). SCEs are generally more sensitive indicators of genotoxic effects than are structural aberrations, and require less effort to analyse. Like structural aberrations, the existence of threshold effects for SCE induction is possible but difficult to prove or disprove, especially without an understanding of the mechanism of formation.

A variation of the basic method involves quantifying cells with a high frequency of SCEs (Moore & Carrano, 1984). In this approach, the distribution of SCEs per cell is determined for each individual, and cells exceeding an arbitrarily chosen number of exchanges (typically those in the upper 5% of the range) are identified as 'high frequency cells' (HFCs). The number of HFCs per subject is then compared between groups of individuals using non-parametric statistics. This method may be useful for detecting differences between individuals or groups of people when differences in the mean frequencies are small.

Interphase cytogenetic analyses

Interphase cells are routinely obtained during the collection of clinical specimens such as Pap smears, blood and urine samples, and skin biopsies. These clinical samples are evaluated for normal and pathological features and provide valuable information for clinical diagnosis. However, historically these samples have had limited usefulness in identifying genetic changes occurring in the tissues, organs or body fluids from which the sample was obtained. Developments in molecular cytogenetics over the past 10 years are allowing new and important types of cytogenetic information to be obtained for cells and cell types that has not been possible previously. One type of molecular technique which is proving to be particularly valuable for these studies is FISH, which allows information on chromosome number and, to a limited extent, chromosome structure to be simply and rapidly obtained from interphase cells (Pinkel, *et al.*, 1986; Gray, *et al.*, 1994; Eastmond & Rupa, 1995).

Interphase studies using FISH focus primarily on aneuploidy, a condition in which the chromosome number of the cell or individual differs from the normal number in that cell type—typically 46 for somatic cells and 23 for germ cells in humans. Aneuploidy in germ cells has been associated with infertility, pregnancy loss, congenital anomalies and mental retardation, whereas in somatic cells this

condition has been linked with cell death and carcinogenesis (Hook, 1985; Oshimura & Barrett, 1986; Hecht & Hecht, 1987). FISH studies for aneuploidy typically utilize DNA probes specific for the centromeric region of one (or occasionally several) chromosome(s) of interest (Eastmond & Rupa, 1995). Following hybridization with the probe, two brightly stained fluorescent hybridization regions are observed at the position of the chromosomal region in both metaphase and interphase somatic cells. Aneuploid cells are identified by counting the number of fluorescent spots in each nucleus. Cells are considered to be aneuploid when the number of hybridization signals differs from the expected two per chromosome in the nuclei of somatic cells and one in the nuclei of germ cells. For technical reasons, this assay is more sensitive for detecting gains in chromosome number (hyperploidy) than for detecting chromosome losses (hypoploidy) (Eastmond & Pinkel, 1990).

In-vitro studies have shown that this FISH assay is effective in detecting hyperdiploidy induced by aneuploidy-inducing agents (Eastmond & Pinkel, 1990). To date, only a limited number of studies using FISH to detect aneuploidy in human populations have been conducted. Some initial studies, primarily in sperm and exfoliated cells, have used probes in an attempt to establish the levels of hyperploidy in the cells of individuals without known exposure to genotoxic agents (e.g. Moore *et al.*, 1993; Robbins *et al.*, 1993; Bischoff *et al.*, 1994). These have been followed by a number of studies (e.g. Robbins *et al.*, 1994; Rupa *et al.*, 1995; Smith *et al.*, 1995) in which FISH, using centromeric DNA probes, has been employed to detect increases in aneuploidy in individuals exposed to benzene, pesticides and chemotherapeutic drugs. In most of these recently published or soon-to-be published studies, a significant increase in hyperploidy was seen in the exposed individuals when compared with control individuals. Although additional studies will be required to validate this assay thoroughly, the results of these initial studies indicate the promise and feasibility of this technique in human biomonitoring.

Since aneuploid cells are identified by counting the number of hybridization regions in the nucleus, hybridization artefacts and other technical problems may significantly affect this assay (Eastmond *et al.*, 1995). For example, the detection of chromosome loss is influenced by the superimposition of signals, associations of centromeric regions in interphase cells and inefficient probe penetration. It is also likely that other cellular and hybridization phenomena can significantly affect the number of cells scored as hyperdiploid. For example, breakage occurring within the chromosomal region targeted by the DNA probe can be incorrectly identified as an additional chromosome (Eastmond, *et al.*, 1994; Rupa, *et al.*, 1995). Although this illustrates a potential limitation of this FISH assay, D.A. Eastmond and co-workers have shown that, by modifying the standard single probe hybridization assay through the use of multicolour FISH with two adjacent probes, this approach can also be used to detect chromosomal breakage in interphase somatic and germ cells (Eastmond *et al.*, 1994; Rupa, *et al.* 1995; D.S. Rupa, unpublished data). Recent studies have shown that this tandem FISH approach is effective for detecting breakage in cells exposed *in vitro* to genotoxic agents, as well as in workers occupationally exposed to pesticides (Rupa, *et al.*, 1995).

One of the advantages of these interphase FISH assays is that they can be performed on most cell types, allowing information on chromosomal alterations to be obtained from the target organs of carcinogens or germ-cell-damaging agents. For example, by assaying cells exfoliated in the urine, it may be possible to measure damage occurring in the bladder of workers or patients exposed to agents that induce bladder cancer. Studies have been initiated that have shown that interphase FISH can be performed on buccal mucosal cells, urothelial cells and sperm, as well as blood cells such as granulocytes and lymphocytes. In addition, it is likely that other cells isolated through skin biopsy, and cells isolated during bronchoalveolar lung lavages or Pap smears could also be assayed. For most of the cell types tested to date, relatively simple preparation procedures can be used. For example, standard cytogenetic fixation procedures can be used to prepare lymphocyte and granulocyte slides. Interphase analyses for aneuploidy and breakage can also be performed on blood smears following simple fixation procedures. However, this approach is not recommended as a routine method, as relatively few nucleated cells are recovered on a standard blood smear and scoring is tedious. It is likely that with relatively minor modifications, such as the use of hypotonic KCl to lyse

the red blood cells in the blood sample, nucleated cells can be concentrated to facilitate scoring as well as to improve probe penetration. This should allow biomonitoring to be performed on minimal amounts of blood, such as that obtained by pricking a finger, and would eliminate the requirement for cell culturing. However, in some cases cell culture may facilitate the detection of chromosome aberrations such as aneuploidy, as cells must pass through a mitosis in order for numerical alterations to occur. Recent studies by Rupa *et al.* (1996) have shown that culturing is not required for the detection of chromosomal breakage in interphase cells. In these studies, similar frequencies of breakage were seen in cultured and non-cultured lymphocytes, and in granulocytes irradiated *in vitro*.

Future cytogenetic techniques

The use of biomarkers for assessing cytogenetic damage has seen tremendous advances in recent years, the most notable of which is FISH. Early efforts were limited to a single colour and repetitive sequences, while the current state of the art involves multiple colours used in combination to identify a wide range of targets comprised of unique and repetitive DNA sequences. The sensitivity and specificity of the various FISH assays have increased accordingly, but there is still a need for additional fluorochromes in the visible range. Ideally, at least 24 unique colours are needed, one for each of the human chromosome types. The actual number of fluorochromes needed to accomplish this is considerably less than 24. For example, two colours are now commonly mixed to yield a third colour. Three fluorochromes can be mixed to yield seven colours (three individually, plus three unique pairs, plus all three together). In general, n fluorochromes can theoretically be mixed to yield $2n - 1$ colours, assuming that the fluorochromes are mixed in equal ratios. Alterations of the ratios will lead to additional colours. As many as 12 colours have now been detected simultaneously (Dauwerse *et al.*, 1992), but the use of this number of colours simultaneously is far from routine. Regular application of large numbers of probes in different colours would significantly increase the frequency of the detectable events, including structural aberrations, chromosomal contents of micronuclei and aneuploidy. Similarly, the application of brighter colours would enable (1) enhanced detection of bound probe, (2) visualization of probes hybridized to smaller targets, (3) reduction of signal fading, and (4) reduced need for computerized image processing, which is expensive and beyond what most laboratories can afford.

The field of interphase cytogenetics has seen significant advances in recent years, also due to applications of FISH. The detection of structural aberrations in interphase with closely spaced probes is significant. Extension of this basic method to other chromosomes and in other species could lead to significant insights into the frequency and distribution of chromosome damage in cell types that are not currently accessible. Such probes would not necessarily have to be targeted to repetitive sequences, as it is now possible to produce unique-sequence probes to specific chromosomal regions by microdissection (e.g., Yokoyama & Sakuragawa, 1995).

It is also possible to perform FISH on tissue slices of archived pathological specimens (Sauter *et al.*, 1995). Application of these methods to epidemiological applications would create new avenues of research in which samples gathered many years previously would become amenable to sensitive and highly specific molecular analyses. Retrospective case–control studies could be performed with relative ease on existing samples. Prospective cohort and nested case–control studies could be planned because the sample storage requirements are straightforward. The relatively small amounts of material that are typically required for molecular analyses could mean that numerous assays might be performed on only a few tissue slices, which would be a distinct advantage for very small or rare samples.

Automation of cytogenetic assays has been a goal for many years. Some progress has been made, for example in the area of karyotyping and aberration detection (e.g. Lundsteen & Piper, 1989; Mayall *et al.*, 1990; Huber *et al.*, 1995). However, fully automatic analysis of metaphase chromosomes is yet to be realized. Automatic metaphase finding shows significant promise for enhancing throughput, and consequently for increasing the sensitivity of aberration detection (e.g. Piper *et al.*, 1994; Vrolijk *et al.*, 1994). Some progress has also been made towards automatic detection of micronuclei in erythrocytes (e.g. Zetterberg & Grawe, 1993) and nucleated cells

(e.g., Vral *et al.*, 1994), but except for isolated instances, such methodology is not used on a routine basis. Automated systems for spot counting in interphase cells would be useful for detection of aneuploidy. In general, automation is expected to permit the analysis of more cells than would otherwise be possible, with the result that the sensitivity for performing biological dosimetry will be increased.

Improved techniques for hybridization are needed, particularly for slides that have not been stored in N_2 gas at –20°C. While some success at hybridizing archived material has been achieved, routine FISH on old material is often difficult. A method that consistently results in high-quality FISH signals would eliminate the need for slide storage at –20°C, and would open up new avenues of research on existing material. For prospective epidemiology studies, cryopreservation of white blood cells offers an attractive alternative to real-time biomonitoring of at-risk populations, since cells can be stored long-term and evaluated as needed in the future. When frozen for viability, lymphocytes may be readily thawed and cultured (McFee *et al.*, 1995) for studies of structural chromosome aberrations, micronuclei or SCEs. After culture, slides may be prepared for FISH immediately or cells may be stored in fixative at –20°C for periods of at least a year. Similarly, interphase analyses using centromere, chromosome or DNA probes can be performed on lymphocytes or granulocytes from selected persons within cohorts. Such approaches may be useful in case–control or case–cohort studies of populations with specific exposures at some time in the past.

One goal of all these methods is to improve the sensitivity and specificity of cytogenetic analyses, so that their utility as biomarkers of exposure and effect can be improved. Developments such as new DNA probes and probe combinations, enhancements in the area of automation and the hybridization of archived slides, as well as improvements in the number and intensity of fluorochromes, are all expected to increase the rate of analysis and to improve the sensitivity and specificity of detection. The result will be that lower levels of exposure can be detected with increased confidence and reliability in populations that could not previously be evaluated.

Special considerations on the use of cytogenetic end-points in epidemiology studies

Epidemiology studies involving cytogenetic endpoints have generally required advanced planning because of the need to culture cells to bring them to metaphase. This is true for studies seeking to measure structural chromosome aberrations, whether by FISH or by classical methods, as well as for studies involving SCEs. This limitation severely restricts the use of archived material for these endpoints. A further limitation for studies using FISH is the requirement for microscope slides to be frozen in nitrogen gas to preserve their ability to hybridize. The reasons for this are not well understood and might be overcome by additional research. An alternative approach to freezing slides is to freeze cells in a manner that maintains their viability. With this approach, selected samples can be thawed and analysed at a later date. Disadvantages of this method include the expense involved in cryopreservation and diminished cell viability upon thawing. For all of these reasons, metaphase-based analyses are most commonly used for case–control or cross-sectional studies, and not for prospective cohort or retrospective studies.

The advent of interphase-based analyses as described in this paper appears to have significant potential for many types of epidemiology studies. Prospective cohort and nested case–control studies could take advantage of these methods for several reasons. The absence of a requirement for tissue culture greatly simplifies the work needed to store samples, although freezing of the microscope slides is still required for FISH. Interphase analyses might be possible on lymphocytes obtained from gravity sedimentation of whole blood cells ('buffy coats'). These cells could be smeared on microscope slides and stored for later analysis. The simplicity of this method, the inherent low cost and the small amount of material required would make this approach applicable to a broad range of studies, including those that require very large sample sizes.

Finally, a comment on inter-reader variation of cytogenetic analyses is in order. Epidemiologists often express concern about the inherent subjectivity of cytogenetic analyses. Any method that requires human judgment as part of the analytical process has some potential for confounding due to interindividual differences in decision-making. One of the significant advantages of molecular

cytogenetic methods, compared to the older, classical methods, is that the end-points of interest are more easily detected and therefore less subjective in their interpretation. For example, a colour 'junction' which is evidence of a chromosome translocation in painted cells is much more obvious than the equivalent event seen in banded preparations. This increase in objectivity should diminish concerns inherent in any large study where there is more than one person who performs the routine analyses and where unavoidable turnover in the technical staff occurs.

In summary, compared with classical cytogenetic approaches, molecular cytogenetic biomarkers offer a number of significant advantages that make them appealing for use in a wide variety of epidemiological studies.

Acknowledgements

This work was performed in part under the auspices of the US DOE by the LLNL under contract No. W-7405-ENG-48 and supported by NIH grant CA59431 (JDT). LGL was supported by NCI interagency agreement Y01-CP3-0573 and DOE/ORAU contract DE-AC05-76OR00033. DAE was supported by a US EPA grant (R 820994-01-1).

References

Awa, A.A. (1974) Cytogenetic and oncogenic effects of the ionizing radiations of the atomic bombs. In: German, J., ed., *Chromosomes and Cancer*, New York, John Wiley, pp. 637–674

Bender, M.A., Awa, A.A., Brooks, A.L., Evans, H.J., Groer, P.G., Littlefield, L.G., Pereira, C., Preston, R.J. & Wachholz, B.W. (1988) Current status of cytogenetic procedures to detect and quantify previous exposures to radiation. *Mutation Res.*, 196, 103–159

Bischoff, F.Z., Nguyen, D.D., Burt, K.J. & Shaffer, L.G. (1994) Estimates of aneuploidy using multicolor fluorescence in situ hybridization on human sperm. *Cytogenet. Cell Genet.*, 66, 237–243

Bogen, K.T. (1993) Reassessment of human peripheral T-lymphocyte lifespan deduced from cytogenetic and cytotoxic effects of radiation. *Int. J. Radiat. Biol.*, 64, 195–204

Bonassi, S., Abbondandolo, A., Camurri, L., Dal Pra, L., De Ferrari, M., Degrassi, F., Forni, A., Lamberti, L., Lando, C., Padovani, P., Sbrana, I., Vecchio, D. & Puntoni, R. (1995) Are chromosome aberrations in circulating lymphocytes predictive of future cancer onset in humans? Preliminary results of an Italian cohort study. *Cancer Genet. Cytogenet.*, 79, 133–135

Brandriff, B.F., Gordon, L.A., Moore, D., II & Carrano, A.V. (1988) An analysis of structural aberrations in human sperm chromosomes. *Cytogenet. Cell Genet.*, 47, 29–36

Brogger, A., Hagmar, L., Hansteen, I.L., Heim, S., Högstedt, B., Knudsen, L., Lambert, B., Kinnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerfving, S. & Sorsa, M. (1990) An inter-Nordic prospective study on cytogenetic endpoints and cancer risk. *Cancer Genet. Cytogenet.*, 45, 85–92

Buckton, K.E. (1976) Identification with G and R banding of the position of breakage points induced in human chromosomes by in vitro X-irradiation. *Int. J. Radiat. Biol.*, 29, 475–488

Buckton, K.E., Jacobs, P.A., Court-Brown, W.M. & Doll, R. (1962) A study of the chromosome damage persisting after X-ray therapy for ankylosing spondylitis. *Lancet*, II, 676–682

Catalan, J., Autio, K., Wessman, M., Lindholm, C., Knuutila, S., Sorsa, M. & Norppa, H. (1995) Age-associated micronuclei containing centromeres and the X chromosome in lymphocytes of women. *Cytogenet. Cell Genet.*, 68, 11–6

Cleary, M.L. (1991) Oncogenic conversion of transcription factors by chromosomal translocations. *Cell*, 66, 619–622

Dauwerse, J.G., Wiegant, J., Raap, A.K., Breuning, M.H. & Van Omen, G.J.B. (1992) Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. *Human Mol. Genet.*, 1, 593–598

Eastmond, D.A. & Tucker, J.D. (1989) Identification of aneuploidy-inducing agents using cytokinesis-blocked human lymphocytes and an antikinetochore antibody. *Environ. Mol. Mutagen.*, 13, 34–43

Eastmond, D.A. & Pinkel, D. (1990) Detection of aneuploidy and aneuploidy-inducing agents in human lymphocytes using fluorescence in situ hydridization with chromosome-specific DNA probes. *Mutation Res.*, 234, 303–318

Eastmond, D.A., Rupa, D.S. & Hasegawa, L.S. (1994) Detection of hyperdiploidy and chromosome breakage in interphase human lymphocytes following exposure to the benzene metabolite hydroquinone using multicolor fluorescence in situ hybridization with DNA probes. *Mutation Res.*, 322, 9–20

Eastmond, D.A. & Rupa, D.S. (1995) Fluorescence in situ hybridization: application to environmental mutagenesis. In: Phillips, D.H. & Venitt, S., eds, *Environmental*

Mutagenesis, Bios Scientific Publications, Oxford, pp. 261–290

Eastmond, D.A., Schuler, M. & Rupa, D.S. (1995) Advantages and limitations of using fluorescence in situ hybridization for the detection of aneuploidy in interphase human cells. *Mutation Res.*, 348, 153–162

Galloway, S., Deasy, D., Bean, C., Kraynak, R., Armstrong, M. & Bradley, O. (1987) Effects of high osmotic strength on chromosome aberrations, sister chromatid exchanges and DNA strand breaks, and the relation to toxicity. *Mutation Res.*, 189, 15–25

Gray, J.W., Pinkel, D. & Brown, J.M. (1994) Fluorescence in situ hybridization in cancer and radiation biology. *Radiat. Res.*, 137, 275–89

Hagmar, L., Brogger, L., Hansteen, I.L., Heim, S., Hogstedt, B., Knudsen, L., Lambert, B., Linnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerfving, S. & Sorsa, M. (1994) Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic Study Group on the Health Risk of Chromosome Damage. *Cancer Res.*, 54, 2919–2922

Hando, J.C., Nath, J. & Tucker, J.D. (1994) Sex chromosomes, micronuclei and aging in women. *Chromosoma*, 103, 186–192

Hecht, F. & Hecht, B.K. (1987) Aneuploidy in humans: dimensions, demography, and dangers of abnormal numbers of chromosomes. In: Vig, B.K. & Sandberg, A.A., eds, *Aneuploidy Part A: Incidence and Etiology*, New York, Alan R. Liss, Inc, pp. 9–49

Heddle, J.A., Hite, M., Kirkhart, B., Movournin, K., MacGregor, J.T., Newell, G.W. & Salamone, M.F. (1983) The induction of micronuclei as a measure of genotoxicity. A report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutation Res.*, 123, 61–118

Heim, S. & Mitelman, F. (1987) *Cancer Cytogenetics*, New York, Alan R. Liss

Hook, E.B. (1985) The impact of aneuploidy upon public health: mortality and morbidity associated with human chromosome abnormalities. In: Dellarco, V., Voytek, P.E. & Hollaender, A., eds, *Aneuploidy*, New York, Plenum Press, pp. 7–33

Huber, R., Kulka, U., Lorch, T.H., Braselmann, H. & Bauchinger, M. (1995) Automated metaphase finding: an assessment of the efficiency of the METAFER2 system in a routine mutagenicity assay. *Mutation Res.*, 334, 97–102

Ishihara, T. & Kumatori, T. (1969) Cytogenetic studies on fishermen exposed to fallout radiation in 1954. *Jpn. J. Genet.*, 44 (Suppl. 1), 242–251

Kleinerman, R.A., Littlefield, L.G., Tarone, R.E., Machado, S.G., Blettner, M., Peters, L.J. & Boice, J.D., Jr (1989) Chromosome aberrations in peripheral lymphocytes and radiation dose to active bone marrow in patients treated for cancer of the cervix. *Radiation Res.*, 119, 176–190

Kleinerman, R.A., Littlefield, L.G., Tarone, R.E., Sayer, A.M., Cookfair, D.L., Wactawski-Wende, J., Inskip, P.D., Block, A.M.W., Ramesh, K.H. & Boice, J.D., Jr (1994) Chromosome aberrations in lymphocytes from women irradiated for benign and malignant gynecological disease. *Radiation Res.*, 139, 40–46

Littlefield, L.G. & Joiner, E.E. (1978) Cytogenetic follow-up studies in six radiation accident victims (16 & 17 years post-exposure). In: *Late Effects of Ionizing Radiation*, Vienna, International Atomic Energy Agency, pp. 297–308

Littlefield, L.G., Kleinerman, R.A., Sayer, A.M., Tarone, R. & Boice, J.D., Jr (1991) Chromosome aberrations in lymphocytes—Biomonitors of radiation exposure. In: Gledhill, B.L. & Mauro, F., eds, *New Horizons in Biological Dosimetry*, New York, Wiley-Liss, pp. 387–397

Lucas, J.N., Tenjin, T., Straume, T., Pinkel, D., Moore, D., II, Litt, M. & Gray, J.W. (1989) Rapid human chromosome aberration analysis using fluorescence in situ hybridization. *Int. J. Radiat. Biol.*, 56, 35–44 (Erratum: *Int. J. Radiat. Biol.*, 56, 201)

Lucas, J.N., Awa, A.A., Straume, T., Poggensee, M., Kodama, Y., Nakano, M., Ohtaki, K., Weier, H.-U., Pinkel, D., Gray, J. & Littlefield, L.G. (1992) Rapid translocation frequency analysis in humans decades after exposure to ionizing radiation. *Int. J. Radiat. Biol.*, 62, 53–63

Lundsteen, C. & Piper, J., eds. (1989) *Automation of Cytogenetics*, Berlin, Springer-Verlag

McFee, A.F., Sayer, A.M. & Littlefield, L.G. (1995) Cryopreservation of lymphocytes for cytogenetic analysis. I. Prompt separation is needed. *Environ. Mol. Mutagen*, 25 (Suppl. 25), 34

Mayall, B.H., Tucker, J.D., Christensen, M.L., van Vliet, L.J. & Young, I.T. (1990) Experience with Athena, a semi-automated karyotyping system. *Cytometry*, 11, 59–72

Moore, D.H., II & Carrano, A.V. (1984) Statistical analysis of high SCE frequency cells in human lymphocytes. In: Tice, R.R. & Hollaender, A., eds, *Basic Life Sciences* Vol. 29, New York, Plenum, pp. 469–479

Moore, L.E., Titenko-Holland, N. & Smith, M.T. (1993) Use of fluorescence in situ hybridization to detect chromosome-specific changes in exfoliated human bladder and oral mucosa cells. *Environ. Mol. Mutagen.*, 22, 130–137

Natarajan, A.T., Vyas, R.C., Wiegant, J.C. & Curado, M.P. (1991) A cytogenetic follow-up study of the victims of a

radiation accident in Goiania (Brazil). *Mutation Res.*, 247, 103–111

Nath, J., Tucker, J.D. & Hando, J.C. (1995) Y chromosome aneuploidy, micronuclei, kinetochores and aging in men. *Chromosoma*, 103, 725–731

Ohtaki, K., Shimba, H., Awa, A.A. & Sofuni, T. (1982) Comparison of type and frequency of chromosome aberrations by conventional and G-staining methods in Hiroshima atomic bomb survivors. *J. Radiat. Res.*, 23, 441–449

Oshimura, M. & Barrett, J.C. (1986) Chemically induced aneuploidy in mammalian cells: mechanisms and biological significance in cancer. *Environ. Mutagen.*, 8l, 129–159

Pinkel, D., Straume, T. & Gray, J.W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl Acad. Sci.*, 83, 2934–2938

Piper, J., Poggensee, M., Hill, W., Jensen, R., Ji, L., Poole, I., Stark, M. & Sudar, D. (1994) Automatic fluorescence metaphase finder speeds translocation scoring in FISH painted chromosomes. *Cytometry*, 16, 17–16

Ramsey, M.J., Moore, D.H., II, Briner, J.F., Lee, D.A., Olsen, L.A., Senft, J.R. & Tucker, J.D. (1995) The effects of age and lifestyle factors on the accumulation of cytogenetic damage as measured by chromosome painting. *Mutat. Res.*, 338, 95–106

Robbins, W.A., Segraves, R., Pinkel, D. & Wyrobek, A.J. (1993) Detection of aneuploid human sperm by fluorescence in situ hybridization: evidence for a donor difference in frequency of sperm disomic for chromosomes 1 and Y. *Am. J. Hum. Genet.*, 52, 799–807

Robbins, W.A., Cassel, M.J., Blakey, D.H., Meistrich, M.L. & Wyrobek, A.J. (1994) Induction of aneuploidy in the sperm of Hodgkin's disease patients treated with NOVP chemotherapy (detection by multi-chromosome fluorescence in situ hybridization). *Environ. Mol. Mutagen.*, 23 (suppl. 23), 58

Rupa, D.S., Hasegawa, L. & Eastmond, D.A. (1995) Detection of chromosomal breakage in the 1cen–1q12 region of interphase human lymphocytes using multicolor fluorescence in situ hybridization with tandem DNA probes. *Cancer Res.*, 55, 640–5

Rupa, D.S., Schuler, M. & Eastmond, D.A. (1996) Detection of aneuploidy and breakage in the 1cen–1q12 region of human lymphocytes induced by genotoxic agents. *Fundamental Appl. Toxicol.*, 30, 46

Salassidis, K., Georgiadou-Schumacher, V., Braselmann, H., Muller, P., Peter, R.U. & Bauchinger, M. (1995) Chromosome painting in healthy irradiated Chernobyl victims: a follow-up study to evaluate the stability of symmetrical translocations and the influence of clonal aberrations for retrospective dose estimation. *Int. J. Radiat. Biol.*, 68, 257–262

Sandberg, A.A. (1990) *The Chromosomes in Human Cancer and Leukemia*, 2nd edn, New York, Elsevier

Sauter, G., Moch, H., Gasser, T.C., Mihatsch, M.J. & Waldman, F.M. (1995) Heterogeneity of chromosome 17 and erbB-2 gene copy number in primary and metastatic bladder cancer. *Cytometry*, 21, 40–46

Schreunemachers, D.M. & Everson, R.B. (1991) Effect of residual splenic function and folate levels on the frequency of micronucleated red blood cells in splenectomized humans. *Mutat. Res.*, 263, 63–67

Smith, M.T., Zhang, L., Rothman, N., Wang, Y., Hayes, R.B., Li, G.L. & Yin, S.N. (1995) Interphase cytogenetics of workers exposed to benzene. *Toxicologist*, 15, 87

Stanbridge, E.J. (1992) Functional evidence for human tumour suppressor genes; chromosome and molecular genetic studies. *Cancer Surv.*, 2, 5–24

Titenko-Holland, N., Moore, L.E. & Smith, M.T. (1994) Measurement and characterization of micronuclei in exfoliated human cells by fluorescence in situ hybridization with a centromeric probe. *Mutat. Res.*, 312, 39–50

Tucker, J.D., Auletta, A., Cimino, M.C., Dearfield, K.L., Jacobson-Kram, D., Tice, R.R. & Carrano, A.V. (1993a) Sister-chromatid exchange: second report of the Genetox program. *Mutat. Res.*, 297, 101–180

Tucker, J.D., Ramsey, M.J., Lee, D.A. & Minkler, J.L. (1993b) Validation of chromosome painting as a biodosimeter in human peripheral lymphocytes following acute exposure to ionizing radiation in vitro. *Int. J. Radiat. Biol.*, 64, 27–37

Tucker, J.D., Lee, D.A., Ramsey, M.J., Briner, J., Olsen, L. & Moore, D.H., II. (1994) On the frequency of chromosomal exchanges in a control population measured by chromosome painting. *Mutat. Res.*, 3131, 93–202

Vral, A., Verhaegen, F., Thierens, H. & de Ridder, L. (1994) The in vitro cytokinesis-block micronucleus assay: a detailed description of an improved slide preparation technique for the automated detection of micronuclei in human lymphocytes. *Mutagenesis*, 9, 439–443

Vrolijk, J., Sloos, W.C., Darroudi, F., Natarajan, A.T. & Tanke, H.J. (1994) A system for fluorescence metaphase finding and scoring of chromosomal translocations visualized by in situ hybridization. *Int. J. Radiat. Biol.*, 66, 287–295

Yokoyama, Y. & Sakuragawa, N. (1995) Improved simple generation of GTG-band specific painting probes. *Cytogenet. Cell Genet.*, 71, 32–36

Zetterberg, G. & Grawe, J. (1993) Flow cytometric analysis of micronucleus induction in mouse erythrocytes by gamma-irradiation at very low dose-rates. *Int. J. Radiat. Biol.*, 64, 555–564

Corresponding author
J.D. Tucker
Biology and Biotechnology Research Program,
PO Box 808, L-452, Lawrence Livermore National
Laboratory, Livermore, CA 94551, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Methodological issues in the use of tumour markers in cancer epidemiology

Z.-F. Zhang, C. Cordon-Cardo, N. Rothman, A.N. Freedman and J.A. Taylor

In this chapter, we review major methodological and practical issues associated with the use of tumour markers. At this stage of development, studies with a combination of tumour, susceptibility and exposure markers are needed to illustrate the link between exposure and biological response and to assess the interactive effects of tumour susceptibility markers in this process. Several practical issues related to the application of tumour markers are discussed, including banking of tumour tissue, setting a laboratory strategy and performing etiological heterogeneity analysis.

Recent advances in technology aimed at identifying genetic mutations and abnormally expressed products have led to an interest in incorporating these alterations as tumour markers in epidemiological studies in an attempt to better understand the natural history and the etiology of cancer. There is a concomitant need for research into methodological issues including study design, statistical analysis, and interpretation as well as practical issues such as tumour tissue storage.

Tumour markers have been defined by Schwartz (1989) as:

> '... substances which can be measured quantitatively by biochemical or immunochemical means in tissue or body fluids to identify the presence of a cancer, possibly the organ where it resides, to establish the extent of tumour burden before treatment as well as to monitor the response to therapy.'

The term has been defined more narrowly as (Klavins, 1989):

> '... substances that are produced by cancer cells and present in the circulation of cancer patients at a higher concentration than in individuals without cancer. It can be used to monitor the course of the disease.'

More broadly, tumour markers can be defined as those abnormal biological products or molecular alterations related to any sequence of multistage carcinogenesis, such as tumour initiation and promotion, and may be measured quantitatively or qualitatively by biochemical, immunochemical, cytogenetic and molecular techniques in human biological materials, including tumour tissues, blood and urine samples, etc. These markers may be employed to predict primary or secondary tumour risk, to establish tumour burden, to subclassify tumours beyond histological classification (in addition to pathological classification), to predict tumour prognosis, to determine treatment strategy and to evaluate chemoprevention or intervention efficacy. They can be employed in primary prevention or etiological research by detecting the relationship between environmental exposure and specific mutations; in secondary prevention studies (i.e. early detection and diagnosis) by identifying a precursor or tumours at the early stage; and in tertiary prevention (i.e. enhanced prognosis) if they predict tumour progression and patient survival.

In this chapter, we will briefly review some tumour markers of relevance to etiological studies of cancer and discuss methodological issues associated with the use of those markers in cancer epidemiology, including issues of study design and statistical analysis. We will also discuss several tumour-marker-related practical issues, including the collection and storage of tumour-marker-related biological materials, the laboratory strategy for measuring tumour markers, and issues associated with the use of tumour markers in prognostic studies.

Etiological heterogeneity

The theoretical framework for the use of tumour markers in cancer epidemiology is based on the hypothesis of etiological heterogeneity. The concept of etiological heterogeneity was first suggested by Kreyberg (1937), and was based on the fact that a relationship might exist between histological types and biological or etiological factors. In epidemiological studies of smoking and lung cancer, Kreyberg (1952, 1954) suggested that lung cancer was histologically and biologically heterogeneous and related to particular exposures. The relationships between stage, grade, histology, anatomic location of the tumour and etiological factors have been explored for a number of different tumour types (Palli *et al.*, 1992; Sturgeon *et al.*, 1994; Vaughan *et al.*, 1995). The study of etiological heterogeneity using tumour markers represents a progression from studies of the relationship of environmental exposures at the anatomical and morphological levels to studies at molecular and genetic levels.

In perhaps one of the earliest studies of this kind to examine etiological heterogeneity at the molecular level, Taylor *et al.* (1992) conducted a study of the relationship between occupational exposures and the activation of the *ras* oncogene in the etiology of acute myeloid leukaemia (AML), employing a conventional control group identified by random digit dialling. In this study, in addition to calculating unadjusted and adjusted odds ratios characterizing the association between occupational factors and AML, the investigators also assessed the association between the occupational risk factors and mutations of the *ras* gene in cases with AML. Several recent studies have now expanded this paradigm by evaluating the etiological heterogeneity of tumours, exploring the relationship between environmental exposures and *TP53* mutations/*p53* overexpression in tumour samples from a variety of tumour types, including lung cancer (Kondo *et al.*, 1992; Suzuki *et al.*, 1992; Taylor *et al.*, 1994), head and neck cancer (Field *et al.*, 1991), oesophageal cancer (Hollstein *et al.*, 1991a), bladder cancer (Spruck *et al.*, 1993; Zhang *et al.*, 1994b), colorectal cancer (Zhang *et al.*, 1995b; Freedman *et al.*, 1996a, 1996b), prostate cancer (Zhang *et al.*, 1994a), stomach cancer (Zhang *et al.*, 1995a), liver cancer (Bressac *et al.*, 1991; Hsu *et al.*, 1991; Ozturk 1991) and skin cancer (Brash *et al.*, 1991). The study of the etiological heterogeneity of tumours at the molecular level may provide great insight into the mechanisms and causal pathways to carcinogenesis, which may lead to appropriate preventive strategies to reduce the incidence of cancer.

Tumour markers

Tumour markers include all of the biological products related to the development and progression of neoplastic disease. Several markers are quantitatively measured and may indicate the burden or extent of the cancer, such as serum carcinoembryonic antigen (CEA), alphafetoprotein (AFP) and prostate-specific antigen (PSA). Since those markers have been discussed frequently in the literature (Chu, 1987; Sell, 1992; Kramer & Srivastava, 1994), we will not discuss them in this chapter. In the following section, we will briefly review three types of tumour markers with etiological implications: cytogenetic markers such as chromosome aberrations, oncogenes such as the *ras* family, and tumour suppressor genes such as the *TP53* gene.

Cytogenetic markers

Since the use of cytogenetic markers has been extensively reviewed by Tucker *et al.* (this volume), we will limit our discussion to the cytogenetic analysis of lymphohaematopoietic malignancies. Cytogenetic markers in this context include chromosome aberrations, sister chromatid exchange and micronuclei, which can be measured in both tumour tissues and peripheral lymphocytes. Specific cytogenetic changes were found to be non-random events involved in multistep carcinogenesis. Similar cytogenetic changes were identified in second primary acute non-lymphocytic leukaemia (ANLL) after chemotherapy (Mitelman *et al.*, 1981; Rowley, 1983) and in patients with occupational exposures (Mitelman *et al.*, 1981; Golomb *et al.*, 1982; Rowley, 1983; Mitelman *et al.*, 1984). A high correlation was observed between sites of cytogenetic aberrations and positions of oncogene or tumour suppressor gene loci (Rowley, 1984).

Environmental exposures, cytogenetic markers and lymphohaematopoietic malignancies. Sandler and Collman (1987) have extensively reviewed issues concerning cytogenetic and environmental factors in the etiology of acute leukaemia. Sandler *et al.*

(1993) investigated the leukaemia risk associated with cigarette smoking and cytogenetic changes in a multicentre case–control study of acute leukaemia in adults. Smoking was found to be more common among patients with specific chromosome abnormalities in acute myeloid leukaemia (AML) [–7 or 7q–, –Y, +13] and in acute lymphocytic leukaemia (ALL) [t(9;22)(q34;q11)]. In a case–case study of 59 patients with newly diagnosed AML (Fagioli *et al.*, 1992), 18 patients had prolonged contact with pesticides and seven patients had exposure to organic solvents. Cytogenetic studies confirmed the frequent occurrence of 5q and/or 7q aberrations in patients with occupational exposure (10 out of 25 cases). These findings revealed that AML in patients occupationally exposed to toxic substances might represent a distinct cytogenetic entity.

In a case–case study of 162 patients with ANLL (Mitelman *et al.*, 1981), 52 patients were occupationally exposed to chemical solvents, insecticides or petrol products, and 110 patients had no history of occupational exposure to potential mutagenic or carcinogenic agents. Clonal chromosomal aberrations were present in 75% of exposed patients, compared with only 32% in the unexposed group. The incidence of certain characteristic karyotypic abnormalities (5q–, 7q–, +8, +21, t(8;21), and t(9;22)) were decidedly more common in exposed than in unexposed patients. In another case–case study of 74 patients with ANLL (Golomb *et al.*, 1982), 25 of the 58 (43%) unexposed patients had a clonal chromosome abnormality, compared with 12 of the 16 (75%) exposed patients ($P = 0.02$). Either a –5/5q– or a –7/7q– was present in 67% of the exposed patients with a chromosome abnormality, compared with 28% of the aneuploid unexposed patients. These studies support the observation that a subset of patients with *de novo* ANLL have a history of occupational exposure and a unique pattern of clonal chromosomal aberrations.

Cytogenetic changes in patients with second acute non-lymphocytic leukaemia after treatment of a primary malignancy. Cytogenetic studies were conducted on 26 patients who developed ANLL or a dysmyelopoietic syndrome after treatment of a primary malignancy (Rowley *et al.*, 1981). Fifteen patients had radiotherapy and chemotherapy, seven had only chemotherapy, and four had only radiotherapy. Twenty-five patients had an abnormal karyotype in myeloid cells. Loss of part or entire chromosomes 5 and/or 7 was noted in 23 of 25 patients with aneuploidy. Loss of chromosome 5 was noted only in patients who previously had malignant lymphoma, whereas loss of chromosome 7 was seen in these patients as well as in those who had other malignancies. Abnormalities of both chromosomes 5 and 7 occurred in 53% of the patients treated with combined therapy and in only 27% of patients treated with either modality alone. Cytogenetic changes were analysed in 76 cases of secondary myelodysplasia (sMDS) and acute non-lymphocytic leukaemia (sANLL) (Johansson *et al.*, 1991). Among the 36 sMDS patients, 23 (64%) displayed clonal chromosomal abnormalities. The most common aberrations were –7, 5q–, –5 and +8. Of the 40 sANLL patients, 30 (75%) cases displayed clonal chromosomal abnormalities. The most frequent aberrations were –7, –5, +8, 7q–, –17 and +21. When the incidences of characteristic cytogenetic abnormalities were correlated with the type of previous therapy, –7 was found to be more frequent in sMDS and sANLL patients who had been exposed to chemotherapy, whereas 5q– was associated with previous exposure to ionizing radiation in sMDS patients. Those results suggest that cytogenetic measures may be employed to predict the risk of second primary tumour after treatment such as radiotherapy and chemotherapy for the first primary tumour.

In summary, many chromosomal abnormalities identified will have diagnostic, prognostic and therapeutic implications. The identification of chromosomal abnormalities directs us to investigate abnormal loci of the genome that harbour the molecular basis responsible for malignant transformation and progression (McClay, 1989). The implementation of interphase cytogenetics by techniques such as fluorescence in-situ hybridization (FISH) will lead to the more frequent use of cytogenetic markers in cancer epidemiological studies.

Oncogenes and tumour suppressor genes

Inherited (germline) or acquired (somatic) gene mutations and altered gene products controlling cell growth, cell death and differentiation are considered to be crucial steps in human carcinogenesis. Molecular studies have defined that aberrations affecting two major types of genes, proto-oncogenes and tumour suppressor genes have a direct role in tumorigenesis and cancer progression.

Oncogenes. Proto-oncogenes are normal cellular genes involved in a wide variety of functions, such as cell growth and signal transduction. The activation of a proto-oncogene into an oncogene yields a gain of function or dominant event. Oncogene activation usually occurs by a somatic mutation, mainly gene amplification or point mutation. These alterations can convert a proto-oncogene from a normal cellular gene to an oncogene and can lead to uncontrolled or neoplastic growth (Taylor, 1989).

Tumour suppressor genes. These genes are also normal cellular genes. Tumour suppressor genes contribute to oncogenicity through their loss of functions, and are considered recessive events. The end result is that the products of these genes are absent or inactivated in the malignant cells. Tumour suppressor gene mutations are the most frequently observed genetic events in cancer. In general, suppressor gene inactivation occurs by a point mutation of one allele and a deletion of the remaining contralateral allele. The loss of both alleles (homozygous deletion) is an alternative but uncommon mechanism. These alterations result in an inability to suppress cell proliferation, and it is followed by tumour development. Since many mutagens are capable of altering tumour suppressor genes, it has been hypothesized that tumour suppressor genes contribute to the development of cancer and may be a critical area in which to study cancer etiology (Hollstein *et al.*, 1991b; Jones *et al.*, 1991; Harris, 1993).

The use of oncogenes and tumour suppressor genes in cancer epidemiology. The use of molecular and genetic alterations of tumour suppressor genes and proto-oncogenes in cancer epidemiology has advanced our understanding of cancer biology and carcinogenesis. Point mutations in tumour suppressor genes (e.g. *TP53*) and oncogenes (e.g. *ras*) may be specific for both tumour type and the critical environmental exposure (etiological heterogeneity). Lung tumours from smokers show a high frequency of G to T transversions in both K-*ras* and *TP53*, and may reflect the molecular fingerprint of carcinogenic constitution of tobacco smoke (Jones *et al.*, 1991). Such molecular epidemiological evidence supports the well established association between smoking and cancer (Vineis & Caporaso, 1995), although it is important to point out that mutational patterns differ for other smoking-related cancers such as bladder cancer.

The wide range of involvement of *TP53* in human tumours and the broad spectrum of mutations make the gene a good candidate for molecular epidemiological studies (Hollstein *et al.*, 1991b; Jones *et al.*, 1991; Harris, 1993; Harris & Hollstein, 1993). *TP53* mutations have been suggested as DNA fingerprints of exposure in a variety of tumours. Dietary AFB_1 exposures are associated with AGG→AGT mutations at codon 249 in liver cancer (Ozturk *et al.*, 1991; Bressac *et al.*, 1991; Hsu *et al.*, 1991). UV exposure may induce CC→TT mutations in skin tumour (Brash *et al.*, 1991). Radon exposures are related to AGG→ATG mutations in codon 249 in lung cancer (Taylor *et al.*, 1994).

Newly developed molecular biological methods, e.g. polymerase chain reaction (PCR), automated sequencing techniques and comparative genomic hybridization (CGH), will accelerate the process of characterizing DNA alterations. The CGH method represents a recently developed molecular cytogenetic screening technique that can be employed to survey entire genomes for variations in DNA sequence copy number, as well as to map chromosome regions with amplifications or deletions in tumour DNA prepared from fresh or archived materials (Kallioniemi *et al.*, 1993; Thompson & Gray, 1993; Houldsworth & Chaganti, 1994; Kallioniemi *et al.*, 1994). CGH is a powerful adjunct to traditional cytogenetic techniques and a useful tool with which to screen for molecular and genetic defects, which will eventually lead to the identification of tumour suppressor genes and oncogenes in solid as well as haematological tumours. By combining advanced methods for characterizing exposure to carcinogens and measuring tumour markers, there is a great potential for further elucidating the etiology of cancers and for the development of strategies for cancer prevention.

Study designs for the use of tumour markers

Tumour marker studies provide some interesting new challenges in study design and statistical analysis. A detailed discussion of study designs for the use of biomarkers in epidemiology is provided by Rothman *et al.* (1995), Hulka & Margolin (1992) and Hulka & Garrett (1996).

Case–case study design

It has been suggested that case–case (or case–series or case-only) studies can be employed to evaluate gene–environment interactions. The critical assumption is that the exposure and genetic factors occur independently and the disease is rare. Under these assumptions, the case–case approach is valid and offers better precision for estimating gene–environment interactions than does the case–control approach (Piegorsch *et al.*, 1994). The case–case study design can be employed to assess the association between exposure and tumour markers and to evaluate etiological heterogeneity between marker-positive and marker-negative tumours (Taylor, 1989; Begg & Zhang, 1994). This study design may be used to evaluate the hypothesis that the two categories of cases, distinguished by the presence or absence of the tumour marker, are characterized by etiological heterogeneity, as evidenced by differences in the strengths of effects of the risk factor in the two case groups. The differences could be due to the fact that the causal pathway differs, or they could merely reflect a different magnitude of effect via the same mechanism. Empirical evidence of such etiological heterogeneity with respect to one or more risk factors would provide strong justification for more detailed investigations of the specific mechanisms of action. This study design consists of a series of incident cases. Ideally, this would be a consecutive series of population-based incident cases. If the ascertainment is not complete, or if, for example, the study is hospital-based, we must assume that case selection for the two disease categories is not influenced differentially by the risk factors associated with case ascertainment.

Suppose that Y is the risk factor of primary interest, assumed for simplicity to be binary, and that W denotes the set of remaining risk factors, where $Y+$ indicates the presence of the risk factor and $Y-$ indicates its absence. Let $X+$ ($X-$) denote the presence (absence) of the tumour marker. Furthermore let $\psi(W)$ be the odds ratio relating Y and X, conditional on W. In the context of our study of *p53* overexpression and bladder cancer (Zhang *et al.*, 1994b), Y represents smoking status, X represents the presence or absence of *TP53* mutations in the tumour samples, and W represents the remaining risk factors. We can evaluate $\psi(W)$ using standard statistical methods such as the Mantel–Haenszel procedure or logistic regression. A test of the null hypothesis that $\psi(W) = 1$ is a test of the hypothesis of etiological heterogeneity, i.e. that the strength of Y as a risk factor is different for the two case groups (e.g. *p53+* and *p53–*).

Case–control study design

In case–control studies, etiological heterogeneity has traditionally been evaluated by two separate analyses: marker-positive cases versus controls and marker-negative cases versus controls. The analytic strategy is to use polychotomous logistic regression (Dubin & Pasternak, 1986). In this model, the relationships between marker-positive cases and controls, and between marker-negative cases and controls are both modelled concurrently using two separate (logistic) regression functions. Let β_1 be the coefficient of the primary risk factor in the logistic regression relating marker-positive cases and controls, and let β_2 be the corresponding parameter relating marker-negative cases and controls. If there are no interactions between Y and W, then β_1 is the conditional log odds ratio of the risk factor on marker-positive disease, and β_2 is the conditional log odds ratio of the risk factor on marker-negative disease. To test the null hypothesis that the two diseases possess etiological heterogeneity with respect to the risk factor, one can test the null hypothesis that $\beta_1 = \beta_2$, i.e. that the two odds ratios are equal. Such a comparison can be accomplished by using, for example, a likelihood ratio test. Quantitative evidence of the degree of departure from the hypothesis can be characterized by the difference in these coefficients, $\beta_1 - \beta_2$. This is the logarithm of the ratio of the two adjusted relative risks of the risk factor, i.e. the relative risk with respect to marker-positive cases and with respect to marker-negative cases, respectively.

Case–case versus case–control study design

It has been shown by our group (Begg & Zhang, 1994) that the odds ratio from the case–case study is theoretically equivalent to the parameter $\beta_1 - \beta_2$ in the polychotomous logistic regression model, and thus evaluation of etiological heterogeneity does not require a conventional control group.

We illustrate the method using data from our own case–case study of the relationship between smoking and *TP53* mutations in patients with bladder cancer (Zhang *et al.*, 1994b). The raw

Table 1. Frequencies by *TP53* mutations and smoking status

Smoking status	Cases		Controls
	p53+	*p53–*	
Smoker	34	43	81
Non-smoker	10	21	64

frequencies are contained in Table 1. For illustrative purposes we have employed a control group consisting of patients with other cancers believed to be unrelated to smoking, although this would not be an ideal control group for a case–control study in general. The odds ratios and confidence intervals are presented in Table 2. The unadjusted odds ratios are calculated directly from the cross-products, as usual, i.e. $\psi = (34 \times 21)/(10 \times 43)$, $\theta_1 = (34 \times 64)/(10 \times 81) = 2.69$, and $\theta_2 = (43 \times 64)/(21 \times 81)$. Note that $\psi = \theta_1/\theta_2$.

We have shown that the odds ratio relating an environmental risk factor to the presence of a biological marker is an appropriate measure for characterizing the degree of etiological heterogeneity between the disease groupings defined by the marker. These observations legitimize the common recent practice of exploring gene–environment associations in case–case studies (Piegorsch *et al.*, 1994). This parameter has been shown to be the ratio of the relative risk of the factor in causing marker-positive disease to the relative risk in causing marker-negative disease. Moreover, it can be estimated directly from an appropriately designed case–case study without the need for a control group (Zhang *et al.*, 1994a, 1994b, 1995a, 1995b; Freedman *et al.*, 1996a, 1996b). The odds ratio obtained from case–case study needs to be interpreted with great caution, since this measure does not indicate the directions of individual relative risk of marker-positive or marker-negative disease. Therefore, the use of a control group is necessary if we wish to estimate the actual relative risks for marker-defined tumour subtypes.

Practical issues in tumour tissue banking

The research strategy to incorporate exposure, susceptibility and tumour markers for a case–control study includes collection and storage of all related biological specimens, such as blood samples (for cases and controls), tumour and normal tissues (for cases only), in addition to the collection of epidemiological exposure data. By collecting and storing blood specimens for all cases and controls, certain tumour susceptibility markers, such as polymorphism in GST-M1, NAT2 and CYP1A1, can be assessed. Other markers such as mutagen sensitivity, DNA repair capacity and haemoglobin/DNA adducts can also be measured. Proper collection and storage of tumour and normal tissues for cases allow the characterization of tumour suppressor genes and oncogenes, as well as the DNA replication repair defect phenotype, and microsatellite instability. By collecting epidemiological data, exposure history can be evaluated and other potential confounding factors can be controlled. Since the issue of blood specimen banking has been discussed by Landi & Caporaso (this volume)

Table 2. Odds ratio estimates

Design	Estimate (95% CI)		
	Parameter	Unadjusted	Adjusted[a]
Case–control	$\theta_1(W)$	2.69 (1.23, 5.87)	3.61 (1.41, 9.29)
	$\theta_2(W)$	1.62 (0.87, 3.00)	2.11 (0.98, 4.54)
	$\theta_1(W)/\theta_2(W) = \psi(W)$	1.66 (0.69, 4.01)	1.71 (0.66, 4.43)
Case–case	$\psi(W)$	1.66 (0.69, 4.01)	1.71 (0.63, 4.66)

[a]Adjusted for age using logistic regression for the case–case approach, and polychotomous logistic regression for the case–control approach.
θ_1 = value positive; θ_2 = value negative.

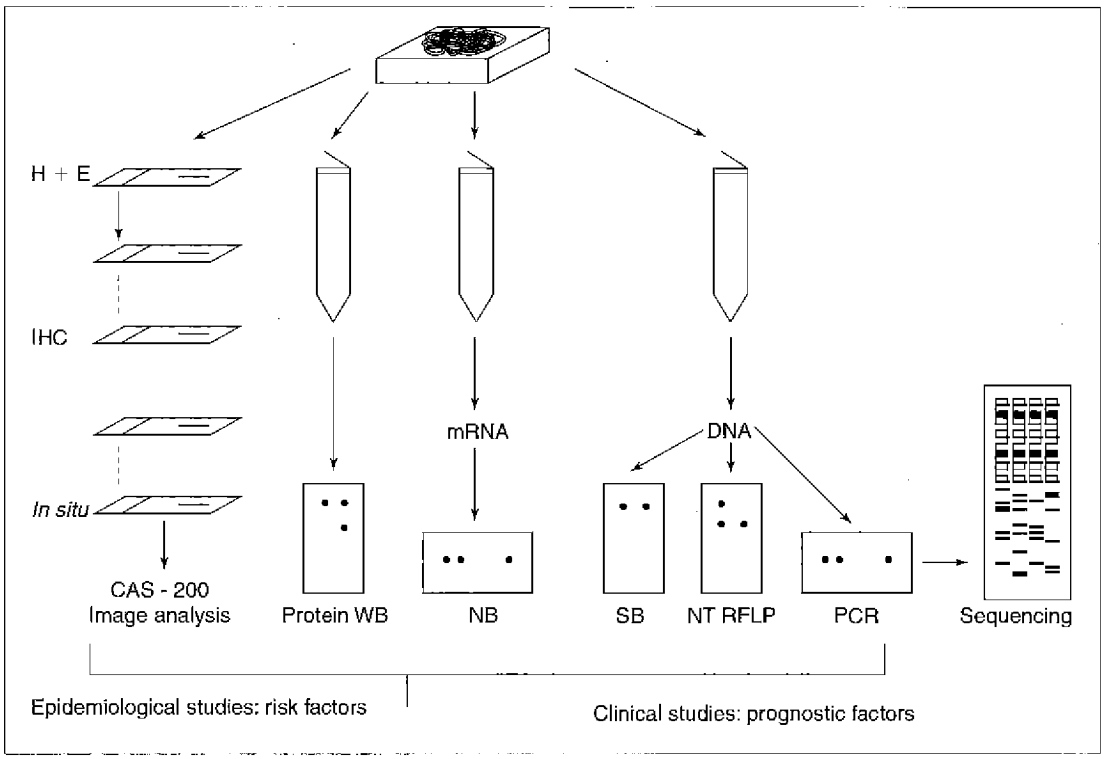

Figure 1. Laboratory strategy for the use of tumour markers.

in this monograph, we will briefly review here issues related to the storage of tumour and normal tissues.

With the increasing demands for tumour tissue samples for biological, clinical or epidemiological studies, tissue banking has become a very important practical issue (Lee *et al.*, 1995). Limited resources, untrained personnel and absence of uniform protocols for tumour tissue banking have created obstacles for the proper and rapid collection, processing and storage of tumour tissues (Grizzle, 1994). In addition, lack of information on diagnostic quality control, histological classification (stage and grade), treatment and outcome may further jeopardize the optimal usage of normal and tumour samples in cancer research.

There is a rapidly increasing ability to evaluate a range of tumour markers in formalin-fixed, paraffin-embedded tissues, such as *p53* nuclear protein accumulation measured by immunohistochemistry (IHC) (Sarkis *et al.*, 1993a, 1993b, 1994, 1995) as well as *TP53* point mutations measured by PCR-SSCP and sequencing techniques (Taylor *et al.*, 1996). The development of assays that can use DNA extracted from formalin-fixed, paraffin-embedded tissue will enhance molecular epidemiological investigation. There are some limitations associated with the use of formalin-fixed, paraffin-embedded materials, mainly due to the fragmentation of DNA. Fresh or frozen samples are preferred, if available, for molecular assays at the DNA or RNA level. Since the storage of formalin-fixed, paraffin-embedded tissues is a standard procedure in medical centres or hospitals, we will focus our discussion on the storage of frozen tissues.

In order to perform epidemiological studies and to share the specimen resource with other investigators in biological and clinical sciences, a centralized programme for research specimen banking is needed in any large research medical centre. Biological specimens can be collected and evaluated for research suitability by trained personnel

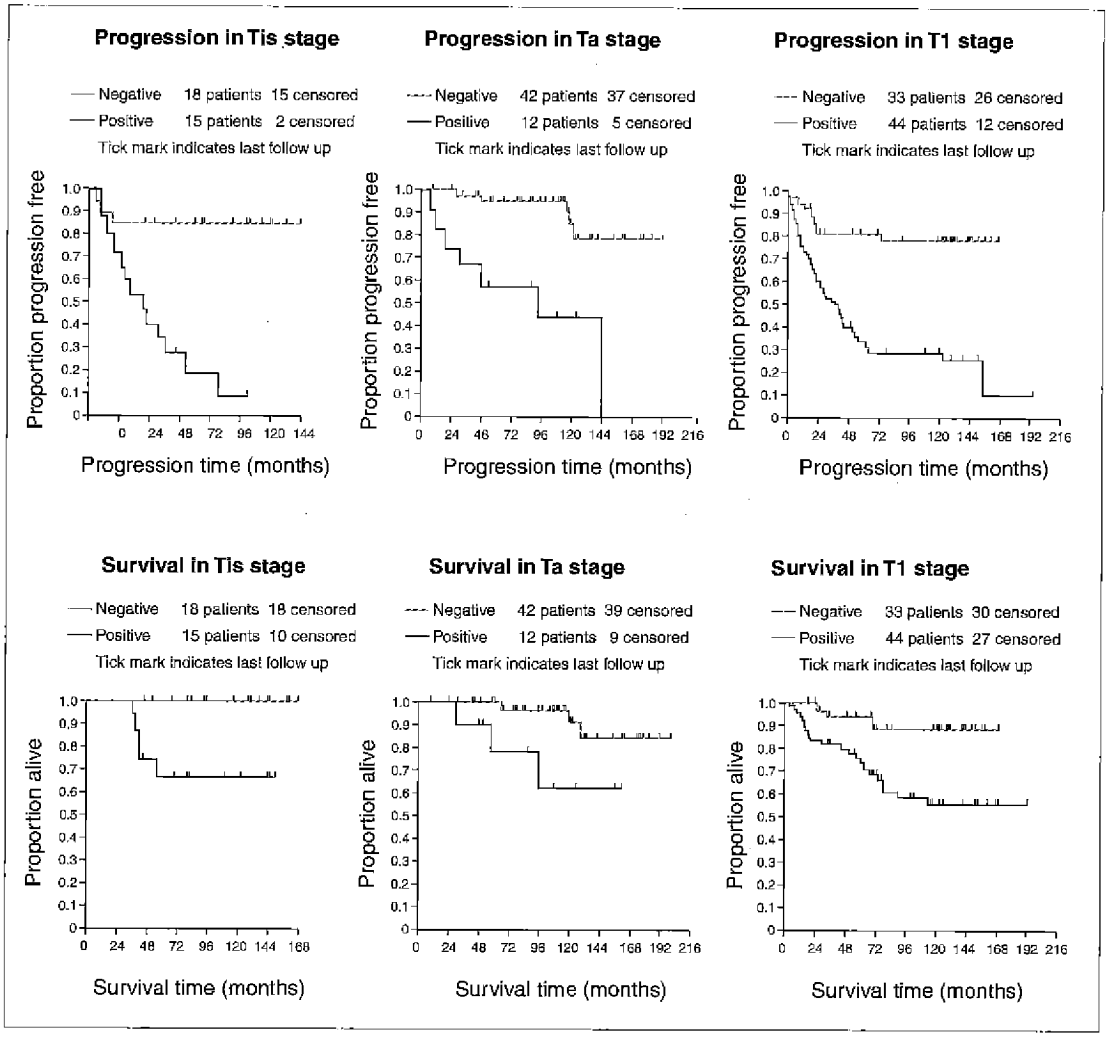

Figure 2. p53 overexpression and prognosis of superficial bladder cancer.

(nurses, technicians and pathologists) and can be distributed among investigators. The materials can be collected and prepared to suit the individual requirements of investigators, and collection and distribution can be centrally documented. The programme can relieve investigators of the burden, in terms of both time and cost, of procuring and preparing a specimen on each occasion. The programme needs to have an effective quality control system in order to evaluate all tissue specimens, assuring investigators of properly diagnosed research material.

At Memorial Sloan Kettering Cancer Center, the tumour banking programme assumes the responsibility of delivering all routine specimens from the operating rooms to pathologists on a regular schedule. This ensures that specimens available for research are as fresh as possible. Pathologists are responsible for determining whether diagnostic requirements have been satisfied before giving the portion of the residual tissue for banking. The specimen is then transferred in an iced specimen container. Normal and tumour tissues are stored in plastic cassettes labelled with a coded number to

protect patients' confidentiality. The tissue is embedded in frozen section support media (OCT) and stored at −70°C. A clinical abstract of the patient's history with an accompanying pathology report will be provided for each specimen only if IRB approval has been obtained to conduct such a study. In general, patients will be informed of the ongoing protocol and sign a informed consent form. A data management system needs to be implemented and a relational database system needs to be established so that tissue bank database can be linked to pathological, epidemiological and clinical data.

Laboratory strategy for the use of tumour markers

A protocol for immunohistochemistry and for genetic analysis of formalin-fixed, paraffin-embedded tissues has kindly been provided by Drs William Bennett and Curtis C. Harris, NCI Laboratory of Human Carcinogenesis (see Appendix). A strategy has been suggested (Cordon-Cardo *et al.*, 1994) whereby, from a single tissue sample, different techniques can be performed to examine immunophenotype and genotype. This strategy is illustrated in Fig. 1. Briefly, using consecutive tissue sections cut at different thicknesses and deposited either on microslides or microtubes, one can: (1) evaluate morphology (i.e. haematoxylin and eosin (H&E) staining), (2) perform immunohistochemistry procedures, and (3) characterize molecular alterations (i.e. Southern blot, restriction fragment length polymorphism (RFLP), and PCR-SSCP and sequencing). Antigen expression and/or its modulation can be analysed by immunohistochemistry on tumour tissue samples. Finally, specimen identification can provide the correlation of laboratory data with epidemiological, pathological and clinical follow-up data (Fig. 1).

Issues concerning the use of tumour markers in prognostic studies

The study of the prognostic value of tumour markers raises several special issues concerning subject selection and adjustment for other known prognostic factors such as stage and grade of the disease. For example, study populations should usually be limited to incident cases diagnosed within a year in order to reduce selection bias from differential survival. Known prognostic factors such as age, sex, grade, stage and treatment need to be controlled in the data analysis when assessing the effect of a tumour marker.

We have conducted a series of studies to evaluate the association between *p53* nuclear overexpression and progression/survival in a group of patients with superficial bladder cancer. *p53* nuclear overexpression was evaluated in tumours of 164 patients (T1 = 77, Ta =54 and Tis = 33) with superficial bladder cancer by immunohistochemistry using the mouse monoclonal antibody PAB1801 on deparaffinized tissue sections. Antibody 1801 detects both wild-type and mutant *p53* proteins. Due to the prolonged half-life of the mutated *p53* products, they accumulate in the nucleus and can usually be detected by immunohistochemical assays. We studied 42 primary bladder tumour tissues to estimate the sensitivity of immunohistochemical (IHC) methods in the prediction of *TP53* mutations (Cordon-Cardo *et al.*, 1994). We found that the highest sensitivity was reached when the cut-off (in terms of percentage of cells with nuclear immunoreactivity) was 20%, and so we have employed 20% as the cut-off point for IHC results in our analysis. The data were first correlated with conventional prognostic parameters, including stage, grade, vascular invasion, age and sex. Various univariate and multivariate analyses were performed. In the study, none of the normal urothelial and stromal cells showed *p53* nuclear overexpression. Patients with bladder tumours were stratified into two groups according to the percentage of cells with nuclear immunoreactivity. Ninety-three patients (56.7%) had none or less than 20% tumour cells with positive nuclear staining (group A), while the remaining 71 (43.3%) had more than 20% tumour cells with nuclear immunoreactivity (group B). Separate analyses of progression and survival were performed for the three stages of superficial bladder cancer (Tis, Ta, T1) after adjusting for age, sex, grade, vascular invasion and adjuvant therapy. Patients in group B had consistent significantly lower progression-free intervals and survival ($P < 0.001$) at all three stages (Fig. 2). These results suggest that superficial bladder cancers exhibiting *p53* nuclear overexpression have a higher rate of disease progression and short survival, and may be useful in selecting appropriate therapy (Sarkis *et al.*, 1993a, 1993b, 1994, 1995).

Appendix. Histology protocol: paraffin sections for immunohistochemical and genetic analyses

1. Background: genetic analysis of archival human tissues

Access to archival human tissues provides many opportunities for both prospective and retrospective studies. Informative assays based on immunohistochemistry and PCR technology are available. Specialized protocols are needed for these studies, and essential elements include confidentiality, proper selection of adhesive coatings and precautions for tissue carry-over in PCR-based studies.

2. Tissue specifications

The optimal tissue sample is a paraffin block containing at least 1 cm^2 of tissue, including both tumour and non-tumour tissue. The non-tumour tissue is used for an internal negative control and for germline analysis. Five-micron sections are used for immunochemistry, and 20-micron sections for microdissection of non-necrotic tissues for genetic analysis.

3. Glass slide specifications: silanation and DEPC treatment

Glass slides must be coated to promote tissue adherence, otherwise tissue sections will be lost during multiple washings and incubations of the immunochemistry protocol. Several coating agents are commonly used, including poly-L-lysine, glue, silane and others; non-biological preparations (i.e. silane) are less commonly contaminated with DNase or RNase than poly-L-lysine and represent a better choice for PCR-based studies. In addition, RNAse precautions must be used for sections intended for microdissection; essential elements include DEPC treatment and baking to inactivate RNase and DNase. Specifications and protocols are listed below:

i. Coated or charged slides. Either commercial or locally prepared slides will be used. Commercially coated slides are available, specify RNase-free.

ii. DEPC treatment protocol. To destroy any RNase or DNase attached to the glass slides, load the slides into a metal rack and place in water containing 0.1% DEPC (diethyl procarbonate) for 15–30 min; wrap the entire rack in aluminum foil and bake at 180°C for at least 2 h. Cool to room temperature and store wrapped in foil at room temperature indefinitely.

Caution: DEPC is a potent protein denaturant and is a suspected carcinogen; it should be handled with care. Wear gloves and safety glasses and work in a chemical fume hood. Point the bottle away from you when opening; internal pressure can lead to splattering.

4. Paraffin section guidelines: RNAse precautions

i. Gloves are worn during microtomy.

ii. Disposable blades are used and replaced between blocks.

iii. The block holder is cleaned with xylene between blocks.

iv. DEPC-treated water is used in the tissue flotation bath.

v. Coated slides (i.e. silane or similar) are prepared with sterile DEPC water and handled only with gloved hands.

vi. Cut 25 sections from each block according to the following specifications. Section nos 1–16 must be 5 microns thick: nos 1 and 16 are stained by haematoxylin and eosin (H&E), nos 2–15 must be unstained. Section nos 17–25 must be 20 microns thick and unstained; put two 20-micron-thick sections on each slide. Coated slides and DEPC water must be used for all sections.

For small tissues, cut only 5-micron sections and place two 5-micron-thick sections on slide nos 17–25.

vii. If the tissue is exhausted during sectioning, make an H&E stain of the last (or close to last) section.

viii. Bake sections at no more than 60°C for no more than 2 h.

ix. Each slide will be labelled with the block number from the paraffin block, the sequential number of the section (i.e. 1–25) and the section thickness (i.e. 5 or 20 microns).

5. Trouble-shooting guide for using this protocol

It is advisable to test the sections produced by a laboratory using this protocol. A common problem is that tissue sections fall off the slide during microwave antigen-retrieval procedures. Usually this is caused by inadequate or improper silanation of sections, although some tissues are more likely than others to fall off the slides (i.e. tissues containing fat or bone). Therefore, investigators are encouraged to send slides from the first 10–20 cases to a histology laboratory, and to request a routine immunostain (i.e. p53 or KI67) using 30 min of microwave antigen retrieval (this is the longest interval commonly used). If tissue sections fall off the slides, then ask the laboratory to remake their silane solutions and re-check their protocol, and then repeat the pilot immunostains until the problem is resolved.

Acknowledgements

The research was supported in part by grants ES-06718 from the National Institute of Environmental Health Services and CA-47538 from the National Cancer Institute, National Institute of Health, Department of Health and Human Services. We are grateful to Drs Colin B. Begg, Barbara S. Hulka, H. Yamasaki and Anthony J. McMichael for their critical review, helpful comments and suggestions, and to Ms Ming Sun for her technical assistance.

References

Begg, C.B. & Zhang, Z.F. (1994) Statistical analysis of molecular epidemiology studies employing case-series. *Cancer Epidemiol. Biomarkers Prev.*, 3, 173–175

Brash, D.E., Rudolph, J.A., Simon, J.A., Lin, A., McKenna, G.J., Boden, H.P., Halperin, A.J. & Ponten, J. (1991) A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl Acad. Sci.*, 88, 10 124–10 128

Bressac, B., Kew, M., Wands, J. & Ozturk, M. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature*, 350, 429–431

Chu, T.M. (1987) Biochemical markers for human cancer. In: Seifert, G., ed., *Morphological Tumor Markers*, New York, Springer-Verlag, pp. 19–46

Cordon-Cardo, C., Dalbagni, G., Saez, G.T., Oliva, M.R., Zhang, Z.F., Rosai, J., Reuter, V.E. & Pellicer, A. (1994) p53 mutations in human bladder cancer: genotypic versus phenotypic patterns. *Int. J. Cancer*, 56, 347–353

Dubin, N. & Pasternak, B.S. (1986) Risk assessment for case-control analysis by polychotamous logistic regression. *Am. J. Epidemiol.*, 123, 1101–1117

Fagioli, F., Cuneo, A., Piva, N., Carli, M.G., Previati, R., Balboni, M., Tomasi, P, Cariani, D., Scapoli, G. & Castoldi, G. (1992) Distinct cytogenetic and clinicopathologic features in acute myeloid leukemia after occupational exposure to pesticides and organic solvents. *Cancer*, 70, 77–85

Field, J.K., Spandidos, D.A., Malliri, A., Gosney, J.R., Yiagnisis, M. & Stell, P.M. (1991) Elevated P53 expression correlates with a history of heavy smoking in squamous cell carcinoma of the head and neck. *Br. J. Cancer*, 64, 573–577

Freedman, A.N., Michalek, A.M., Marshall, J.R., Mettlin, C.J., Petrelli, N., Black, J., Zhang, Z.F., Satchidanand, S.K.S. & Asirwatham, J.E. (1996a) The relationship between smoking and p53 overexpression in colorectal cancer. *Br. J. Cancer*, 73, 902–908

Freedman, A.N., Michalek, A.M., Marshall, J.R., Mettlin, C.J., Petrelli, N., Black, J., Zhang, Z.F., Satchidanand, S.K.S. & Asirwatham, J.E. (1996b) Familial and nutritional risk factors for p53 overexpression in colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.*, 5, 285–291

Golomb, H.M., Alimena, G., Rowley, J.D., Vardiman, J.W., Testa, J.R. & Sovik, C. (1982) Correlation of occupation and karyotype in adults with acute nonlymphocytic leukemia. *Blood*, 60, 404–411

Grizzle, W.E. (1994) Tissue resources in the detection and evaluation of markers. In: Srivastava, S., Lippman, S.M., Hong, W.K. & Mulshine, J.L., eds, *Early Detection of Cancer: Molecular Markers*, Armonk, NY, Futura Publishing Company, Inc., pp. 9–76

Harris, C.C. (1993) p53: At the crossroads of molecular carcinogenesis and risk assessment. *Science* 262, 1980–1981

Harris, C.C. & Hollstein, M.C. (1993) Clinical implications of the p53 tumor-suppressor gene. *N. J. Engl. Med.*, 329, 1318–1327

Hollstein, M.C., Peri, L., Mandard, A.M., Welsh, J.A., Montesano, R., Metcalf, R.A., Bak, M. & Harris, C.C. (1991a) Genetic analysis of human esophageal tumors from two high incidence geographic areas: frequent p53 base substitutions and absence of ras mutations. *Cancer Res.*, 51, 4102–4106

Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C.C. (1991b) p53 mutations in human cancers. *Science*, 253, 49–53

Houldsworth, J. & Chaganti, R.S. (1994) Comparative genomic hybridization: an overview [comment]. *Am. J. Pathol.*, 145, 1253–1260

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J. & Harris, C.C. (1991) Mutational hotspot in the p53 gene in human heptocellular carcinoma. *Nature*, 350, 427–428

Hulka, B.S. & Garrett, P. (1996) Influence of biomarkers on the design, methods, and interpretation of epidemiologic studies. *J. Toxicol. Environ. Health*, 40, 389

Hulka, B.S. & Margolin, B.H. (1992) Methodological issues in epidemiologic studies using biologic markers. *Am. J. Epidemiol.*, 135, 200–209

Johansson, B., Mertens, F., Heim, S., Kristoffersson, U. & Mitelman, F. (1991) Cytogenetics of secondary myelodysplasia (sMDS) and acute nonlymphocytic leukemia (sANLL). *Eur. J. Haematol.*, 47, 17–27

Jones, P.A., Buckley, J.D., Henderson, B.E., Ross, R.K. & Pike, M.C. (1991) From gene to carcinogen: a rapidly evolving field in molecular epidemiology. *Cancer Res.*, 51, 3617–3620

Kallioniemi, O.P., Kallioniemi, A., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F. & Pinkel, D. (1993) Comparative

genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Seminars Cancer Biol.*, 4, 41–46

Kallioniemi, O.P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F.M., Gray, J.W. & Pinkel, D. (1994) Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes, Chromosomes & Cancer*, 10, 231–243

Klavins, J.V. (1989) Search for specific tumor markers. In: Ting, S.W., Chen, J.S. & Schwartz, M.K., eds, *Human Tumor Markers*. New York, Elsevier. pp. 101–104

Kondo, K., Umemoto, A., Akimoto, S., Uyama, T., Hayashi, K., Ohnishi, Y. & Monden, Y. (1992) Mutations in the P53 tumor suppressor gene in primary lung cancer in Japan. *Biochem. Biophys. Res. Commun.*, 183, 1139–1146

Kramer, B.S. & Srivastava, S. (1994) Molecular biology in lung cancer prevention: from the bench to the bedside. In: Srivastava, S., Lippman, S.M., Hong, W.K. & Mulshine, J.L., eds, *Early Detection of Cancer, Molecular Markers*, Armonk, NY, Futura Publishing Company, Inc., pp. 1–12

Kreyberg, L. (1937) Lungecancer. Discussion (Norw.) Detnorske medicinske Selskab. *Norske Mag. f. Lagev*, 98, 36–41

Kreyberg, L. (1952) One hundred consecutive primary epithelial lung tumors. *Br. J. Cancer*, 6, 112–119

Kreyberg, L. (1954) The significance of histological typing in the study of the epidemiology of primary epithelial lung tumors. A study of 466 cases. *Br. J. Cancer*, 8, 199–208

Lee, L.W., Griffith, J., Zenick, H. & Hulka, B.S. (1995) Human tissue monitoring and specimen banking: opportunities for exposure assessment, risk assessment, and epidemiologic research. *Environ. Health Perspect.*, 103 (Suppl. 3), 3–8

McClay, E.F. (1989) Epidemiology of bone and soft tissue sarcomas. *Seminars Oncol.*, 16, 264–272

Mitelman, F., Nilsson, P.G., Brandt, L., Alimena, G., Gastaldi, R. & Dallapiccola, B. (1981) Chromosome pattern, occupation, and clinical features in patients with acute nonlymphocytic leukemia. *Cancer Genet. Cytogenet.*, 4, 197–214

Mitelman, F., Nilsson, P.G. & Brandt, L. (1984) Fourth International Workshop on Chromosomes in Leukemia 1982: Correlation of karyotype and occupational exposure to potential mutagenic/carcinogenic agents in acute nonlymphocytic leukemia. *Cancer Genet. Cytogenet.*, 11, 326–331

Ozturk, M. (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*, 338, 1356–1359

Palli, D., Bianchi, S., Decarli, A., Cipriani, F., Avellini, C., Cocco, P., Falcini, F., Puntoni, R., Russo, A. & Vindigni, C. (1992) A case–control study of cancers of the gastric cardia in Italy. *Br. J Cancer*, 65, 263–266

Piegorsch, W.W., Weinberg, C.R. & Taylor, J.A. (1994) Non-hierarchical logistic models and case-only design for assessing susceptibility in population-based case–control studies. *Stat. Med.*, 13, 153–162

Rothman, N., Stewart, W.F. & Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: A matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev*, 4, 301–311

Rowley, J.D. (1983) Chromosome changes in leukemic cells as indicators of mutagenic exposures. In: Rowley, J.D. & Ultmann, J.E., eds, *Chromosomes and Cancer*, New York, Academic Press. pp. 139–157

Rowley, J.D. (1984) Biological implications of consistent chromosome rearrangements in leukemia and lymphoma. *Cancer Res*, 44, 3159–3168

Rowley, J.D., Golomb, H.M. & Vardiman, J.W. (1981) Nonrandom chromosome abnormalities in acute leukemia and dysmyelopoietic syndromes in patients with previously treated malignant disease. *Blood*, 58, 759–767

Sandler, D.P. & Collman, G.W. (1987) Cytogenetic and environmental factors in the etiology of the acute leukemia in adults. *Am. J. Epidemiol.*, 126, 1017–1032

Sandler, D.P., Shore, D.L., Anderson, J.R., Davey, F.R., Arthur, D., Mayer, R.J., Silver, R.T., Weiss, R.B., Moore, J.O., Schiffer, C.A. *et al.* (1993) Cigarette smoking and risk of acute leukemia: associations with morphology and cytogenetic abnormalities in bone marrow. *J. Natl Cancer Inst.* 85, 1994–2003

Sarkis, A.S., Dalbagni, G., Cordon-Cardo, C., Zhang, Z.F., Sheinfeld, J., Fair, W.R., Herr, H.W. & Reuter, V.E. (1993a) Nuclear overexpression of p53 protein in transitional cell bladder carcinoma: a marker for disease progression. *J. Natl Cancer Inst.*, 85, 53–59

Sarkis, A.S., Zhang, Z.F., Cordon-Cardo, C., Melamed, J., Dalbagni, G., Sheinfeld, J., Fair, W.R., Herr, H.W. & Reuter, V.E. (1993b) P53 nuclear overexpression in Ta non-invasive papillary transitional cell carcinoma of the bladder. *Int. J. Oncol.*, 3, 355–360

Sarkis, A.S., Dalbagni, G., Cordon-Cardo, C., Melamed, J., Zhang, Z.F., Sheinfeld, J., Fair, W.R. & Herr, H.W. (1994) Association of p53 nuclear overexpression and tumor progression in carcinoma *in situ* of the bladder. *J. Urol.*, 152, 388–392

Sarkis, A.S., Bajorin, D.F., Reuter, V.E., Herr, H.W., Netto, G., Zhang, Z.F., Schultz, P.K., Cordon-Cardo, C. & Scher,

H.I. (1995) Prognostic value of p53 nuclear overexpression in patients with invasive bladder cancer treated with neoadjuvant MVAC. *J. Clin. Oncol.*, 13, 1384–1390

Schwartz, M.K. (1989) Tumor markers in diagnosis and screening. In: Ting, S.W., Chen, J.S. & Schwartz, M.K., eds, *Human Tumor Markers*, New York, Elsevier, pp. 3–16

Sell, S. (1992) Cancer markers of the 1990s. In: Sell, S., ed., *Serological Cancer Markers*, New Jersey, Humana Press, pp. 1–17

Spruck, C.H.I., Rideout, W.M., Olumi, A.F., Ohneseit, P.F., Yang, A.S., Tsai, Y.C., Nichols, P.w., Horn, T., Hermann, G.G., Steven, K., Ross, P.K., Yu, M.C. & Jones, P.A. (1993) Distinct pattern of p53 mutations in bladder cancer: relationship to tobacco usage. *Cancer Res.*, 53, 1162–1166. (Erratum: *Cancer Res.* 1993, 53 (Suppl. 10), 2427)

Sturgeon, S.R., Hartge, P., Silverman, D.T., Kantor, A.F., Linehan, W.M., Lynch, C. & Hoover, R.N. (1994) Association between bladder cancer risk factors and tumor stage and grade at diagnosis. *Epidemiology*, 5, 218–225

Suzuki, H., Takahashi, T., Kuroishi, T., Suyama, M., Ariyoshi, Y. & Ueda, R. (1992) p53 mutations in non-small cell lung cancer in Japan: association between mutations and smoking. *Cancer Res.*, 52, 734–736

Taylor, J.A. (1989) Oncogenes and their applications in epidemiologic studies. *Am. J. Epidemiol.*, 130, 6–13

Taylor, J.A., Sandler, D.P., Bloomfield, C.D., Shore, D.L., Ball, E.D., Neubauer, A., McIntyre, O.R. & Liu, E. (1992) ras oncogene activation and occupational exposure in acute myeloid leukemia. *J. Natl Cancer Inst.*, 84, 1626–1632

Taylor, J.A., Watson, M.A., Devereux, T.R., Michels, R.Y., Saccomanno, G. & Anderson, M. (1994) p53 mutation hotspot in radon-associated lung cancer [see comments]. *Lancet*, 343, 86–87

Taylor, J.A., Li, Y., He, M., Mason, T., Mettlin, C., Vogler, W.J., Maygarden, S. & Liu, E. (1996) p53 mutations in bladder tumors from arylamine-exposed workers. *Cancer Res.*, 55, 294–298

Thompson, C.T. & Gray, J.W. (1993) Cytogenetic profiling using fluorescence in situ hybridization (FISH) and comparative genomic hybridization (CGH). *J. Cell. Biochem.*, 17G, 139–143

Vaughan, T.L., Davis, S., Kristal, A. & Thomas, D.B. (1995) Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: adenocarcinoma versus squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, 4, 85–92

Vineis, P. & Caporaso, N. (1995) Tobacco and cancer: epidemiology and the laboratory. *Environ. Health Perspect.*, 103, 156–160

Zhang, Z.F., Aprikian, A., Sarkis, A.S., Zeng, Z.S., Pollack, D., Cordon-Cardo, C., Fair, W.R. & Begg, C.B. (1994a) Factors associated with p53 accumulation in prostatic adenocarcinoma. *Int. J. Oncol.*, 4, 897–901

Zhang, Z.F., Sarkis, A.S., Cordon-Cardo, C., Dalbagni, G., Melamed, J., Aprikian, A., Pollack, D., Sheinfeld, J., Herr, H.W., Fair, W.R., Reuter, V.E. & Begg, C.B. (1994b) Tobacco smoking, occupation, and p53 nuclear overexpression in early stage bladder cancer. *Cancer Epidemiol., Biomarkers Prev.*, 3, 19–24

Zhang, Z.F., Karpeh, M.S., Lauwers, G.Y., Marrero, A.M., Pollack, D., Cordon-Cardo, C. & Begg, C.B. (1995a) Risk factors and p53 nuclear overexpression in early stage intestinal-type adenocarcinoma of the stomach. *Cancer Detection Prev.*, 19, 156–164

Zhang, Z.F., Zeng, Z.S., Sarkis, A.S., Klimstra, D.S., Charytonowicz, E., Guillem, J.G., Cordon-Cardo, C., Cohen, A.M. & Begg, C. (1995b) Family history of cancer, body weight and p53 nuclear over-expression in Duke's C colorectal cancer. *Br. J. Cancer*, 71, 888–893

Corresponding author

Z.-F. Zhang
Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, Box 44, New York, NY 10021, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Quality control of biomarker measurement in epidemiology

D. Gompertz

The throughput and complexity of a biomarker assay will determine the amount of effort that can be expended on quality control and assurance. Clinical chemistry quality control procedures can be readily applied to simpler chemical analysis such as blood lead and cholesterol, but even complex cell-based biomarker techniques such as *HPRT* mutation analysis and cytogenetics benefit from a formal quality control approach. Collaborative interlaboratory exercises are essential, especially when no certified reference material is available, and these can play a central role in the control of laboratory drift. Recommendations are made for the quality control of biomarker measurement based on clinical chemistry techniques. These include recommendations for coding samples so that the laboratory scientist is unaware of exposure status and for the use of formal laboratory protocols.

Incorporation of biomarker measurements into epidemiological studies can greatly increase their power and the strength of associations that are found. The measurement and control of error in such studies are largely the province of the epidemiologist (White, this volume), however, the quality of the laboratory measurements remains the responsibility of the laboratory scientist providing the biomarker test (Armstrong *et al.*, 1992; Aitio & Apostoli, 1995). The importance of 'transitional' studies in determining intra-individual and interindividual variability and the factors affecting these has been stressed (Hulka, 1991; Schulte & Perera, this volume) but the role of laboratory quality control in minimizing laboratory error needs to be emphasised. Failure to control laboratory drift over time and a lack of comparability with other studies can cause major difficulties in interpretation and analysis in epidemiological studies dependent on biomarker measurements (Vineis *et al.*, 1993).

The range of biomarkers that might be incorporated into an epidemiological study requires a variety of techniques, extending from sophisticated analytical chemistry (atomic absorption spectrometry, isotope dilution mass spectrometry) to more biological methods dependent upon cell culture and molecular biology techniques. The throughput and complexity of a method can vary from the ability to analyse 100 samples a day at a cost of a few dollars per sample (e.g. blood lead measurements) to labour-intensive methods that cost over $1000/sample, with a batch of a dozen taking several days to process (e.g. dioxin and haemoglobin adduct measurements). High-throughput automated analytical techniques are more easily adapted to the quality control techniques that have been so effective in improving clinical chemistry performance; however, low-capacity methods requiring cell culture and other biological approaches are more difficult to control in this way. Quality control procedures developed for the high-throughput methods clearly will not be directly applicable to the more biological assays. There are, however, lessons that can be learnt from the clinical chemistry approach to quality control and assurance that are applicable to all assays, and specific recommendations can be made for biomarker measurements being used in epidemiological studies. In this paper, a general approach to laboratory quality control and assurance is described, followed by examples taken across the range of biomarkers available, i.e. blood lead, serum cholesterol, dioxin, haemoglobin adducts, ^{32}P-postlabelling, *HPRT* mutation assays and cytogenetic scoring.

As clinical chemistry developed, the precision of an assay was monitored during an analytical run by observing the repeatability of duplicate measurements and the repeatability of quality control standards analysed throughout the run. (For

definitions of agreed terms used in laboratory quality assurance programmes, see Aitio & Apostoli, 1995). Day-to-day reproducibility was assessed by repeated analysis of the quality control samples and, at times, carry-over samples from a previous batch. Numerous texts have been written explaining these techniques of quality control (Whitehead, 1977; Aitio & Apostoli, 1995) and the statistical approaches towards managing quality control data, including statistical rules for the acceptability of quality control results (Westgard et al., 1981). These statistical approaches have been used to establish the amount of effort required to monitor precision throughout a batch, i.e. the number of quality control replicates that are needed. This number depends on the length of a run that can be expected without significant drift, and as automated clinical chemistry instruments become more sophisticated, less attention to within-run drift is required. Some analyses such as blood lead measurement (see below) still require frequent quality control samples to allow for changing instrumental conditions. Strict protocols and automatic data handling can ensure tight control of day-to-day quality performance. Assay performance can be compared with similar assays using coefficients of variation (CV) (within and between batches); these are measures of analytical performance and do not reflect intra-individual or interindividual variation.

These quality control techniques can demonstrate control of 'imprecision' (reproducibility) within any batch of samples and on a day-to-day basis, but do not give any indication of the 'trueness' (or accuracy) of the result. This can only be obtained by reference to 'definitive' or 'reference' methods or by use of certified reference material (Aitio & Apostoli, 1995). However, for newer assays and more biological assays, such materials and methods are unlikely to be available and the only approach available to the laboratory scientist is to establish comparability with other laboratories. Interlaboratory assessment exercises have played a central role in establishing the performance of the different analytical methods described below, and for some analyses there are ongoing interlaboratory quality assurance schemes on a national or international level.

Cross-sectional epidemiological studies may, in practice, take considerable time to perform and thus longitudinal control of laboratory performance can be as important as in cohort studies involving multiple measurements on an individual at different time periods. The control of long-term laboratory drift is a complex issue (Broughton *et al.*, 1986). A simple approach is to establish a pool of quality control samples at the outset, but the usefulness of this approach depends on the long-term stability of these samples; it may be difficult to distinguish between laboratory drift and gradual deterioration of quality control material. The number of quality control analyses performed over the course of a study may not necessarily be large enough to establish statistically the extent of laboratory drift (Broughton *et al.*, 1986). However, the regular participation in interlaboratory quality assurance exercises combined with the use of in-house drift controls will give the laboratory scientist and the collaborating epidemiologists some reassurance.

There are special problems associated with longer-term epidemiological studies, during which there may be advances in analytical techniques or changes in instrumentation, reagents or scientific staff. It is the responsibility of the analyst, when change is foreseeable, to establish the comparability of the old and new techniques in terms of sensitivity (analytical), precision and bias. Over the short term, new and old methods should be run in parallel using the same pool of quality control material or samples. Methods that depend on visual scoring (colonies, chromosomal aberrations) may require considerable training and inter-reader comparison periods to ensure that changing the technician does not produce significant shifts in background or sensitivity. Interlaboratory comparisons before and after any change are desirable to ensure that any change in bias not detected by in-house control methods is found and documented. There is evidence from a number of fields that long-term laboratory performance is best ensured by adherence to an agreed and detailed protocol and to the rule that individual analysts are not permitted to make modifications without an assessment of the likely effects on the performance characteristics of the assay. There is always the temptation for an originator of a new assay to incorporate improvements as experience is gained, but this can cause major difficulties during a study and its analysis.

In clinical laboratory science, it is accepted that quality assurance extends beyond analytical quality

control to selection of sampling materials (syringes, tubes), maintaining sample integrity during transport and before analysis, control of contamination, sample handing and documentation. Quality control of this pre-analytical phase can be monitored by the field survey staff submitting duplicate 'blind' samples from the same individual. It is now generally accepted that all 'exposed' and 'control' samples from epidemiological studies should be coded and analysed without the laboratory scientist knowing their status. It is the joint responsibility of the epidemiologist and the laboratory scientist to establish a coding system that is effective and allows reliable tracking of a sample from receipt to reporting.

Protocols for the collection and documentation of samples should be agreed between the laboratory scientist and epidemiologist. The timing of the sample in relation to recent exposures may be critical for environmental contaminants with short elimination half-lives and should be agreed as the protocol is developed, but it should always be recorded. Similarly the time between collection and storage should be known. The number and origin of each batch of sample containers and a selection from each batch should be kept for future reference; this is especially important if there is the possibility of contamination of the containers before use.

Lead

Blood lead is an example of a biomarker analysis that has been used successfully in the workplace for research and regulation and in population studies. The analysis can use automated sampling equipment and has a reasonable throughput (50–100 samples/day), and so classical clinical chemistry quality control and assurance techniques can be used. Intra-batch precision (a typical CV of 3–5%) is usually monitored by duplicate analysis of all the samples with quality control material analysed after every six to eight actual samples. Inter-laboratory quality assurance exercises are essential for maintaining laboratory performance. In the United Kingdom (UK) scheme, a blind 'spiked' sample is sent to contributing laboratories every 2 weeks for analysis. Performance is assessed in comparison with the mean result from the 100+ UK and overseas participants ('consensus mean'). Earlier studies have shown that this consensus mean is a good reflection of the actual amount of lead present as measured by a definitive method such as isotope dilution analysis (Bullock *et al.*, 1986). Stability of quality control material on long-term storage and the continuing availability of interlaboratory comparison has enabled long-term studies (10+ years) of the effect of environmental change on population blood lead levels in the UK to be performed (Delves *et al.*, 1996).

Blood lead is a well-established biomarker; the analytical methods and their quality control have matured over the last decade and so lead can be used as a model for other toxic metal analyses and, to a great extent, for other chemical contaminants.

Cholesterol

Serum cholesterol measurement is one of the usual range of analyses offered by clinical chemistry laboratories with a possible throughput of hundreds of samples per day. It is also part of the lipid biomarker profile that is used in the epidemiology of cardiovascular disease. It is probably this latter use that has been responsible for the attention that has been paid to the control of this measurement on both a short- and long-term basis. In the USA, detailed recommendations have been produced for the measurement of total cholesterol in blood [National Cholesterol Education Program (NCEP), 1988]. More recently Westgard *et al.* (1991) have shown, for these recommendations, the number of control measurements per run required and the analytical coefficient of variation and bias that are needed to produce a laboratory performance deemed acceptable in practice, i.e. an intralaboratory CV of less than 3%. More recently, Bachorik *et al.* (1995) have produced for NCEP recommendations and criteria of analytical performance for measurements of low-density lipoprotein (LDL) cholesterol. These authors discussed the problems of maintaining linkage with existing epidemiological data as newer laboratory methods evolve. They recommended that the most prudent course at that time was to continue to measure LDL-cholesterol by methods similar to those used to establish the epidemiological database on which the relationships between cardiovascular risk and LDL-cholesterol had been established.

In longer-term epidemiological studies, during which there may be considerable changes in population mean levels of the biomarker of interest, it

is important that laboratory drift is identified and controlled. Internal quality control procedures are not necessarily able to detect such drift and some external source of comparison is required. Standard reference sera are available for cholesterol analyses but the availability of a regular proficiency testing scheme, such as the National External Quality Assessment Scheme (NEQAS) in the UK, has the advantage that any laboratory is able to compare its performance with a number of others. For serum cholesterol, the current interlaboratory coefficient of variation in this scheme is typically 3.8% ($n = 438$).

Although cholesterol measurements are used on a daily basis in a clinical setting and in population screening programmes, as well as in large-scale cohort studies, the maintenance of accuracy and precision over a long period still presents difficulties, and considerable care is required to monitor performance using ongoing internal laboratory quality control procedures, external reference material, stored sample pools and external quality assurance schemes (Broughton *et al.*,1986).

Dioxins

Polychlorinated dibenzodioxins (PCDDs) and dibenzofurans (PCDFs) occur as complex mixtures in ultra-trace (picogram) amounts in biological fluids, but in recent years the development of highly sensitive high-resolution gas-chromatography mass spectrometry has allowed them to be quantitated. The methods for measurement involve effective sample clean-up and extraction techniques, are labour intensive and do not have a high throughput (Rappe *et al.*, 1991). The considerable progress that has been made to ensure good quality control of these measurements has depended on the development of detailed in-house protocols (Patterson *et al.*, 1991), the use of certified reference materials (Rymen, 1994) and interlaboratory proficiency testing exercises (Yrjanheikki, 1989; Stephens *et al.*, 1992).

An extensive and impressive interlaboratory exercise was initiated by WHO Europe in 1989 (Stephens *et al.*, 1992). In this study three samples each of human breast milk and blood were sent to 19 participating laboratories from 14 countries. Two of the three samples were spiked with known amounts of PCDDs and PCDFs; the third remained unspiked. None of the laboratories knew the identity of the individual samples. The study design resulted in approximately 4000 individual pieces of PCDD/PCDF data being generated by a variety of analytical methods and by laboratories of differing experience. A number of statistical analyses of the data were performed. The simplest was the comparison of coefficients of variation for intralaboratory and interlaboratory variability. Eleven out of 16 laboratories produced similar values for the various analytes in milk, with average intralaboratory CVs of 20–30%. Nine of 15 laboratories produced data for blood with average intralaboratory CVs less than 30%. Several approaches were used to identify the principal determinants of the quality of data produced. The results showed that the experience of the laboratory was of more importance than the actual method or sophistication of the instrumentation used. The authors concluded that laboratories that had well established quality assurance programmes, which were followed carefully, produced superior results.

The results from this study show that even these complex and demanding assays with low throughput can be subjected to formal quality assurance procedures and they also show the importance of interlaboratory comparison exercises and high quality reference material in obtaining and maintaining consistent results.

Carcinogen–protein adducts

Protein adducts reflect the dose of an active metabolite of the carcinogen produced in a target tissue and are thus useful surrogates for the shorter-lived equivalent DNA species (Skipper & Tannenbaum, 1990; Wild & Pisani, this volume). Carcinogen–protein adducts have been described for ethylene oxide, benzo[*a*]pyrene, 4-aminobiphenyl and aflatoxin B_1, among others. There has been little interlaboratory effort to document analytical performance, although the technology (high-resolution gas-chromatography mass spectrometry) is similar to that used for dioxin analysis.

An interlaboratory study has been organised to investigate performance in the measurement of *N*-(2-hydroxyethyl)valine in human globin, a haemoglobin adduct biomarker of ethylene oxide exposure (Törnqvist *et al.*, 1992). In this study, samples of globin from eight individuals were sent to four laboratories in different countries. The results varied from laboratory to laboratory, and up

to 10-fold in extent for some samples. Some, but not all, of the differences were shown to be the result of different internal standards. However, it appeared that, although the absolute measures could not be relied upon, the ranking of the samples was reasonably consistent.

This interlaboratory study contrasts with the dioxin study reported in the same year by Stephens *et al.* (1992). Dioxin estimations were much more well established, lessons had been learnt from previous round-robin exercises and standard material was becoming available. This adduct study illustrates the problem of establishing interlaboratory studies for new assays when there are only a limited number of laboratories involved worldwide and when it is difficult to produce standards of adequate quality and in reasonable quantity. However, important lessons were learnt from the study, including the lack of suitability of this biomarker at that time for large-scale epidemiology without considerably more development.

^{32}P-postlabelling for DNA adduct analysis

Current ^{32}P-postlabelling techniques for the detection and quantitation of the adducts of carcinogenic compounds with DNA have great sensitivity, being able to detect one benzo[*a*]pyrenediol epoxide residue per 10^7 nucleotides. Although this assay is still in its early stages of development, it has already shown its usefulness in studying exposure to environmental and occupational pollutants (Perera *et al.*, 1992; Nielsen *et al.*, 1996). The assay is complex, with extraction, digestion, labelling, chromatography and counting stages, and individual laboratories have developed their own variants. This complexity and the amount of ^{32}P required per assay limits the throughput possible and thus the ability to incorporate a number of quality control samples in any batch.

An interlaboratory trial has been performed involving 15 laboratories in eight countries to determine the extent of reproducibility of the assay as it was then being performed (Phillips & Castegnaro, 1993). The laboratories were sent four samples of extracted DNA—two were mouse DNA after in-vivo treatment with polycyclic aromatic carcinogens, one sample was from lung tissue of a smoker and one was a control sample. The participants were asked details of their individual protocols. Three methods were used by the laboratories and the between-laboratory coefficients of variation ranged from 56–70% for the 'standard' method to 65–98% for the butanol-enrichment method. These high coefficients of variation are not unusual in any laboratory discipline at the early stages of method development and may reflect the wide range of conditions employed at each stage by the different laboratories. It is also not totally clear whether the adducted samples were completely stable during transit to the participating laboratories. However, the results showed that there was a general qualitative agreement between the laboratories and they also showed the possibilities for improvement in interlaboratory precision. The importance of stable and relevant reference material is once again emphasized.

HPRT mutations in lymphocytes

Considerable effort has been made by more experienced laboratories to establish how *HPRT* mutational frequencies vary from day to day in the same individual, between individuals and between laboratories and with different assay techniques (Cole & Skopek, 1994; Robinson *et al.*, 1994). Thus these laboratories have established much of the 'transitional' information required when deciding how to use this end-point in an epidemiological study, especially the extent of intra-individual and interindividual variability in control and exposed groups.

Two laboratories (Universities of Vermont and Sussex; see Cole & Skopek, 1994) have compared split and repeat samples from the same individual. These laboratories found similar mutation frequencies although there were differences in experimental techniques. Both found that the observed mutant frequency for any individual donor may vary considerably between experimental determinations (two- to eightfold). It may be that differences in laboratory protocols are of less importance than each laboratory establishing and keeping to a well defined protocol throughout a study. Robinson *et al.* (1994) compared data sets from four laboratories and reported protocol variations, some of which may contribute to interlaboratory variation. The examples of protocol variation that might be important were: preincubation of the mononuclear cells in the presence of mitogen before cloning, media, serum, interleukin-2 (IL-2) source and concentration, type of feeder cells and irradiation

level, cell density per well, 6-thioguanine concentration, and technique and level of experience of each laboratory worker. The routine splitting of blood samples or performing a larger experiment also varied between laboratories and was clearly an important decision to make.

Quality control of this assay is clearly difficult, as the throughput does not allow a large number of replicate samples for control purposes. The use of a strict protocol in any one laboratory is of importance, as is the investigation of the performance of new batches of sera, biochemicals and other consumables. Albertini and Hayes (this volume) emphasise that, for this, as for all culture systems, attention to reagents and conditions is critical. The maintenance of a full computerised database giving details of all the experimental variables (including the laboratory workers or combinations of laboratory workers involved, as well as media components such as serum batch) in addition to the raw data (plate counts) is particularly useful (J. Cole, personal communication). Albertini and Hayes (this volume) note that the ability to cryopreserve the mononuclear cell fraction for subsequent testing provides a control for intralaboratory drift and that such standards could become available for interlaboratory comparisons.

Cytogenetics

Although molecular methods for assessing cytogenetic damage are advancing rapidly, chromosome aberration analysis has been used over the last 30 years as a biological measure of radiation exposure (Tucker *et al.*, this volume). There have been several interlaboratory collaborative exercises comparing radiation dose estimation by cytogenetic analysis. These have covered either the whole assay and analysis or just the scoring of aberrations. For example, Lloyd *et al.* (1987) dispatched aliquots of whole irradiated blood to the participating laboratories for culture and analysis, while Garcia *et al.* (1995) sent unstained coded slides to their participants.

All cell culture systems need close attention to be paid to reagents and culture conditions (Albertini & Hayes, this volume), and the maintenance of quality performance depends greatly on adherence to a well defined laboratory protocol. It is well established that the scoring must be of coded samples, with the scorer 'blinded' to the exposure/control status. Experienced laboratories have routines for maintaining scoring standards and training new scorers. One such laboratory keeps a set of coded slides, which have been scored by experienced persons and show good agreement, to train recruits. This laboratory circulates coded slides amongst the experienced scorers and, in experiments, prepares replicate slides from the same culture for scoring by several people (D.C. Lloyd, personal communication).

Quality control of cytogenetic techniques involving cell culture and scoring stages presents different problems from the control of the analysis of a chemical biomarker. Three key recommendations that can be made are as follows: (1) there should be development of a detailed laboratory protocol, (2) there should be a scoring policy which ensures that all slides are scored 'blind' and that consistency is maintained between different scorers by regular exchange of slides, and (3) where possible, comparisons with other laboratories should be established on a regular basis. As newer methods, such as the fluoresence in-situ hybridization (FISH) technique, are incorporated into population studies, laboratory intercomparison exercises will be necessary to establish an external validity.

Discussion

The examples described here show the considerable efforts that laboratory scientists have made to control the intralaboratory and interlaboratory variability of their assay procedure. The quality control techniques used for the simpler and more routine chemical assays (lead and cholesterol) are reflected in low interlaboratory CVs and an increased confidence in their accuracy (trueness). As assay techniques become more sophisticated and more demanding, and as the end-points are more biological, quality control procedures are more difficult to incorporate. However, intralaboratory and interlaboratory comparisons do give confidence that an assay is performing reliably and indicate the extent of its inherent variability.

A lesson that comes from these differing laboratory assays is that best performance is associated with the development of a written laboratory protocol that includes appropriate control procedures. As assays are developed and are then used in transitional studies, a stage occurs when the assay is sufficiently mature for it to be documented in such

a manner. Once this has been achieved, deviations from an agreed protocol are not acceptable without formal investigation of the effects of change on precision and, where possible, accuracy. This discipline is essential for assays that need to be maintained throughout a prospective cohort study with minimum laboratory drift.

The examples that have been discussed here come from a range of laboratory disciplines, but it is clear that they all reflect the need for the analyst/scorer to be unaware of the status (exposed/control) of any sample or of the value assigned to material circulated in interlaboratory comparison studies. It may not be possible to disguise quality control material so that the laboratory scientist is unaware of its origin, but the submission of a limited number of coded replicate samples, which are not identifiable as such, will supplement in-house quality control procedures.

The wide range of biomarkers that may be used in cancer epidemiology makes it difficult to prescribe rules for quality control and assurance in any detail. For some chemical assays, certified reference material is available, while for some of the more biological assays it is difficult to see how such material could be produced. An interlaboratory consensus can, however, often be agreed, using a pool of material, cryopreserved cells or at least pre-prepared slides. It is clear that, although the precision obtained by a laboratory may be good, any laboratory needs an external reference to give it confidence that its measures have a validity and are not drifting over the period of a study.

Conclusions

There are a number of conclusions that emerge from this review of quality control of biomarker assays:

1. For a range of biomarkers, laboratory performance depends on the use of an agreed written protocol that specifies details of reagents, sampling and analytical procedures, calibration and quality control methods.
2. Laboratories should take advantage of interlaboratory comparison exercises, whether the assays are being newly developed or are mature assays in routine use.
3. The laboratory scientist should, as a general rule, be unaware of the exposure/control status of the samples submitted.
4. Incorporation of a biomarker assay into an epidemiological investigation requires close collaboration between the epidemiologist and the laboratory scientist, to take advantage of the techniques that are available for minimizing and documenting laboratory error.

Acknowledgements

I thank those colleagues who have helped me to review quality assurance procedures in their specialist fields: Dr D. Bullock, Dr E. Bruner, Dr J. Cole, Dr H.T. Delves, Dr D.C. Lloyd, Dr D. Shuker, Dr J.R. Startin. I also thank the UK Department of the Environment for support during this study.

References

Aitio, A. & Apostoli, P. (1995) Quality assurance in biomarker measurement. *Toxicol. Lett.*, 77, 195–204

Armstrong, B.K., White, E. & Saracci, R. (1992) Principles of exposure measurement in epidemiology. *Monographs Epidemiol. Biostatist.*, 21, 257–258

Bachorik, P.S. & Ross, J.W., for the National Cholesterol Eduction Program Working group on Lipoprotein Measurement (1995) National cholesterol education program recommendations for measurement of low-density lipoprotein cholesterol: executive summary. *Clin. Chem.*, 41, 1414–1420

Broughton, P.M.G., Holder, R. & Ashby, D. (1986) Long-term trends in biochemical data obtained from two population surveys. *Ann. Clin. Biochem.*, 23, 474–486

Bullock, D.G., Smith, N.J. & Whitehead, T.P. (1986) External quality assessment of assays of lead in blood. *Clin. Chem.*, 32, 1884–1889

Cole, J. & Skopek, T.R. (1994) Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. *Mutat. Res.*, 304, 33–105

Delves, H.T., Diaper, S.J., Oppert, S., Prescott-Clarke, P., Periam, J., Dong, W., Colhoun, H. & Gompertz, D. (1996) Blood lead concentrations in United Kingdom have fallen substantially since 1984. *Br. Med. J.*, 313, 883–884

Garcia, O.F., Ramalho, A.T., Di Giorgio, M., Mit, S.S., Espinoza, M.E., Manzano, J., Nasazzi, N. & Lopez I. (1995) Intercomparison in cytogenetic dosimetry among five laboratories from Latin America. *Mutat. Res.*, 327, 33–39

Hulka, B.S. (1991), Epidemiologic studies using biological markers: issues for epidemiologists. *Cancer Epidemiol. Biomarkers Prev.*, 1, 13–19

Lloyd., D.C., Edwards, A.A., Prosser, J.S., Barjaktarovic, N., Brown, J.K., Horvat, D., Ismail, S.R., Koteles, G.J., Almassy, Z., Krepinsky, A., Kucerova, M., Littlefield, L.G.,

Mukherjee, U., Natarajan, A.T. & Sasaki, M.S. (1987) A collaborative exercise on cytogenetic dosimetry for simulated whole and partial body accidental irradiation. *Mutat. Res.*, 179, 197–208

National Cholesterol Education Program (1988) Current status of blood cholesterol measurement in clinical laboratories in the United States. *Clin. Chem.*, 34, 193–201

Nielsen, P.S., de Pater, N., Okkels, H. & Autrup, H. (1996) Environmental air pollution and DNA adducts in Copenhagen bus drivers—effect of GSTM1 and NAT2 genotypes on adduct levels. *Carcinogenesis*, 17, 1021–1027

Patterson, D.G. Jr, Isaacs, S.G., Alexander, L.R., Turner, W.E., Hampton, L., Bernert, J.T. & Needham, L.L. (1991) Method 6: determination of specific polychlorinated dibenzo-*p*-dioxins and dibenzofurans in blood and adipose tissue by isotope dilution-high resolution mass spectrometry. In: Rappe, C., Buser, H.R., Dodet, B. & O'Neill, I.K., eds, *Environmental Carcinogens, Methods of Analysis and Exposure Measurement*, Vol. 11, *Polychlorinated Dibenzodioxins and Dibenzofurans* (IARC Scientific Publications No. 108), Lyon, IARC, pp. 299–342

Perera, F.P., Hemminki, K., Gryzbowska, E., Motykiewicz, G., Michalska, J., Santella, R.M., Young, T.-L., Dickey, C., Brandt-Rauf, P., DeVivo, I., Blaner, W., Tsia, W.-Y. & Chorazy, M. (1992) Molecular and genetic damage in humans from environmental pollution in Poland. *Nature*, 360, 256–258

Phillips, D.H. & Castegnaro, M. (1993) Results of an interlaboratory trial of ^{32}P-postlabelling. In: Phillips, D.H., Castegnaro, M. & Bartsch, H., eds, *Postlabelling Methods for Detection of DNA Adducts* (IARC Scientific Publications No. 124), Lyon, IARC, pp. 35–49

Rappe, C., Buser, H.R., Dodet, B. & O'Neill, I.K., eds. (1991) *Environmental Carcinogens, Methods of Analysis and Exposure Measurement*, Vol. 11. *Polychlorinated Dibenzodioxins and Dibenzofurans* (IARC Scientific Publications No. 108), Lyon, IARC

Robinson, D.R., Goodal, K., Albertini, R.J., O'Neill, J.P., Finette, B., Sala-Trepat, M., Moustacchi, E., Tates, A.D., Beare, D.M., Green, M.H.L. & Cole, J. (1994) An analysis of in vivo *hrpt* mutant frequency in circulating T-lymphocytes in the normal human population: a comparison of four data sets. *Mutat. Res.*, 313, 227–247

Rymen, T., (1994) History of BCR work on dioxins. *Fresenius J. Anal. Chem.*, 348, 9–22

Skipper, P.L. & Tannenbaum, S.R. (1990) Protein adducts in the molecular dosimetry of chemical carcinogens. *Carcinogenesis*, 11, 507–518

Stephens, R.D., Rappe, C., Hayward, D.G., Nygren, M., Startin, J., Esboll, A., Carle, J. & Yrjanheikki, E.J. (1992) World Health Organization international intercalibration study on dioxins and furans in human milk and blood. *Anal. Chem.*, 64, 3109–3117

Törnqvist, M., Magnusson, A.-L., Farmer, P.B., Tang, Y.-S., Jeffrey, A.M., Wazneh, L., Beulink, G.D.T., van der Waal, H. & van Sittert, N.J. (1992) Ring test for low levels of *N*-(2-hydroxyethyl)valine in human hemoglobin. *Anal. Biochem.*, 203, 357–360

Vineis, P., Schulte, P.A. & Vogt, R.F., Jr (1993) Technical variability in laboratory data. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology—Principles and Practices*. San Diego, Academic Press

Westgard, J.O., Barry, P.L., Hunt, M.R. & Groth, T. (1981) A multi-rule Shewart chart for quality control in clinical chemistry. *Clin. Chem.*, 27, 493–501

Westgard, J.O., Petersen, P.H. & Wiebe, D.A. (1991) Laboratory process specifications for assuring quality in the US National Cholesterol Education Program. *Clin. Chem.*, 37, 656–661

Whitehead, T.P. (1977) Quality control in clinical chemistry. New York, John Wiley

Yrjanheikki, E.J. (1989) Levels of PCBs, PCDDs and PCDFs in breast milk. *Environmental Health Series, 34*, World health Organization, Regional Office for Europe. Copenhagen, FADL Publishers

D. Gompertz
MRC Institute for Environment and Health,
University of Leicester,
Leicester LE1 9HN,
UK

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Sample collection, processing and storage

M.T. Landi and N. Caporaso

We review issues related to the inclusion of biospecimens in epidemiological studies. Technical advances and the revolution in molecular biology have rendered the use of biomarkers increasingly feasible in epidemiological investigations, however the cost and complexity require interdisciplinary expertise and careful attention to methodological detail in order to ensure validity. The widespread banking of biospecimens for long-term (cohort) studies requires special attention to be paid to these issues. Blood, urine and tumour tissue are in common use in medicine and at least some aspects of sample handling derives directly from this clinical experience, although special considerations apply in the epidemiological setting. An increasingly broad array of biospecimen types have been studied, including exhaled air, nail clippings, buccal cells, saliva, semen, faeces and breast milk. Relevant issues in the processing, storage, shipping, timing of collection and safety procedures are examined in terms of their potential to distort results. The role of carefully developed quality control protocols is emphasized. In order to take full advantage of the opportunities afforded by the use of biomarkers in epidemiological studies, careful attention to biospecimen processing, the stability of the biomarker and the precautions to be taken during transportation and storage of samples is necessary.

The appropriate collection, processing and storage of biospecimens form an essential but often overlooked component to any study that includes biomarkers. If the type or the quantity of the biological substrate is not adequate for the marker to be analysed, or if it is less than optimally processed at any stage in the sequence between collection (handling, labelling, processing, aliquoting, storage) and assay, analyses and subsequent epidemiological inference will suffer. Contamination, volatilization (e.g. methylethylketone), the short half-life of some substances (e.g. nicotine) and the potential that the collection procedure itself influences the substance to be measured (e.g. chromium levels from stainless steel needles used for venepuncture) are design issues that must be anticipated in advance (Bernard, 1995). Since inaccuracy or imprecision can result in increased variation, misclassification and loss of power, this is a major concern to epidemiologists. In case-control studies, if case samples are handled differently from control samples, differential misclassification may occur. These issues take on special importance in cohort studies, where large numbers of samples and the time and resource investment warrant careful planning to ensure efficient, safe and focused procedures (Kaaks *et al.*, 1994)

It is critical to collect information linked to samples in order to interpret the markers optimally: time and date of collection, recent diet and supplement use, reproductive information (i.e. menstrual cycle), recent smoking, current medication use, recent medical illness, storage conditions, etc. can be crucial for later interpretation of results. For example, information on current medication use may be crucial to the interpretation of a metabolic phenotype, i.e. quinidine use and the debrisoquine phenotype.

Overall quality assurance involves the systematic application of optimum procedures to ensure valid, reproducible and accurate results. The most fundamental component is probably the adoption of standardized operating procedures for each aspect of biospecimen handling. As part of a quality assurance plan, stored specimens could be tested on a regular basis to detect sample deterioration. Another aspect of quality assurance involves aliquoting material into multiple small

vials and storing each person's specimen in at least two different physical locations to reduce the likelihood of loss of a large volume of specimen as a result of accidental thawing due to freezer failure.

When samples are used later, they should be selected from specimens that received the same treatment throughout the storage process or the same variations in handling; this should be controlled for in the statistical analysis.

Careful records of disbursements are also needed, i.e. which specimens, how much material remains, and documentation on factors such as thawing which could influence future uses of the material. Automatic edit checks (which can only be overridden by specific authorization of the principal investigator) may be used to block disbursement of the last specimen from an individual.

In this work, we highlight some relevant issues regarding the collection of biological specimens in epidemiological studies. The work does not attempt to provide an encyclopaedic review of all markers, specimen types or collection techniques, but rather attempts to identify some central issues along with examples.

Types of biospecimens

Various types of biological materials can be collected for epidemiological studies, depending on the study design, the markers of interest and the availability of an assay. With the rapid development of molecular biology techniques, it has become critical to collect samples, planning not only for the main biomarkers of interest, but also to process and store material in a way that allows for new biomarkers to be tested in the future.

Several prospective studies have considered blood specimens stored at low temperature in biological banks as a source of information. Biological markers from samples collected before the onset of disease and stored until clinical expression may provide essential information on exposure to endogenous and exogenous factors not biased by the metabolic effects of illness (Winn *et al.*, 1990). Other biological specimen banks have been established to store specimens from persons who have already developed a disease, to characterize the history of the disease. Many collections of pathological material exist, and have increasing value with the availability of improved techniques to extract and characterize DNA from tumour and normal tissue. The usefulness of these collections is greatly enhanced if there are associated epidemiological data. For example, a series of lung tumour specimens would be of much greater interest if age, gender, smoking, occupational exposure and clinical/survival information were also readily available.

In general, biospecimens that can be collected safely with little discomfort to the donor hold the greatest promise for use in large-scale studies. A list of biospecimens suitable for epidemiological studies follows. This list, although far from being exhaustive, can provide an insight into adequate procedures for sample collection, processing and storage.

Blood

Collection of blood provides material that can give physiopathological and genetic information, can reflect the biological rhythms (circadian, menstrual, etc.) of the host, and can reveal recent or remote exposures or their sequelae. Detailed descriptions of techniques relevant to clinical studies are available in clinical laboratory manuals and so this treatment will only highlight major issues of concern to the epidemiologist. The use of skilled technicians and precise procedures when performing phlebotomy are important because painful, prolonged or repeated attempts at venepuncture can cause patient discomfort or injury and result in less than optimum quality or quantity of sample. As an example, prolonged venepuncture can induce the release of prolactin, or increase white blood cells (Statland *et al.*, 1978). Obtaining blood through too narrow a needle can result in haemolysis, with distortion of cell counts and electrolytes (especially potassium). The donor's position while providing the specimen may also influence analyte levels; e.g. serum cholesterol may be higher in standing subjects compared with supine subjects because of orthostatic decreases in plasma volume (Kieldsen *et al.*, 1983). Prolonged use of a tourniquet (i.e. greater than 1 min) can result in haemoconcentration. Any of these factors can result in pre-analytic variation in measurements. Thus, careful review of general and specific factors that may impact on a specific assay is mandatory before study initiation.

Blood is a great source of materials that can be collected for different purposes. Plasma/serum,

lymphocytes, monocytes, erythrocytes, granulocytes and platelets can be obtained through venepuncture and appropriate separation of blood components.

To avoid multiple venepunctures when a test requires a large blood sample, an evacuated tube system with interchangeable glass tubes can be used. Evacuated tubes are commercially prepared with or without additives and with sufficient vacuum to draw a predetermined blood volume (2–20 ml per tube).The different tubes are generally indicated by their colour-coded stoppers:

- Red-top tubes contain no additives. These tubes are used for tests performed on serum samples.
- Lavender-top tubes contain EDTA, commonly used clinically for complete blood counts (CBC).
- Green-top tubes contain heparin (sodium, lithiu or ammonium).
- Blue-top tubes contain sodium citrate and citric acid. Draw volume may be 2.7 or 4.5 ml.
- Black-top tubes contain sodium oxalate. Draw volume may be 2.7 or 4.5 ml.
- Yellow-top tubes contain acid-citrate-dextrose (ACD) solution. Draw volume is 12 ml.
- Grey-top tubes contain a glycolytic inhibitor (such as sodium fluoride, powdered oxalate or glycolytic/microbial inhibitor). Draw volume may be 3 to 10 ml. These tubes are used most often for glucose determinations in serum or plasma samples.

Specific advantages and disadvantages of various anticoagulants are described below. Nutrients and other substances often require special additives to allow analyses, i.e. metaphosphoric acid to measure vitamin C content.

There are important differences between the different anticoagulants. Heparin acts by binding to antithrombin III and thus accelerates the inactivation of thrombin and other clotting factors. The disadvantages of heparin include the presence of impurities and the non-uniform position of substituents, its ability to bind to many proteins, the potential to bind to platelets and cause agglutination, its potential loss of anticoagulant effect in aged blood, and its influence on T-cell proliferation. Low-molecular-weight heparin avoids some of these problems. EDTA works by calcium chelation and is used clinically in haematology studies, but can be influenced by Mg^{2+} concentration. It is well suited to DNA-based assays but has problems for cytogenetic analyses (increases of sister chromatid exchange, decreases of mitotic index, etc.).

Citrate also works by calcium chelation and is used in coagulation studies and blood banking. It is optimal for assays conducted on lymphocytes and DNA.

There have been a limited number of studies comparing various anticoagulants in investigation settings. For example, a study on the effects of time, temperature and additives on a functional assay of C1 inhibitor (a component of the complement cascade) showed that plasma containing heparin or polybrene interfered with the functional assay; on the other hand, EDTA-treated or citrated plasma and serum kept at room temperature were adequate for the assay (Nielsen *et al.*, 1994).

Although there are anecdotal reports of occasional problems with heparin in PCR assays, studies generally find that there are no major differences in the use of EDTA or heparin. For example, Storch *et al.* (1994) compared the two anticoagulants for inhibitory effects on the detection of cytomegalovirus from washed leukocytes in specimen transport tubes. Evaluation was made by the centrifugation/shell vial culture techniques, the pp65 antigenaemia assay and PCR. For each assay, the results with heparin and EDTA were equivalent.

There are other anticoagulants useful for special applications. If only serum is needed, there is no need for anticoagulants. To reduce contamination, it is best if serum is separated from other blood components as soon as possible.

Many newer methods have been described that allow the collection of small quantities of blood adequate for the characterization of DNA. These methods do not require venepuncture or low-temperature conditions during collection, storage and shipping. Examples of these are dried blood specimens which are derived from whole blood spotted directly—or anticoagulated with EDTA before spotting—onto clean slides and air dried at room temperature. The slides with blood smears can be transported or stored at room temperature, and serve as a good source of high-molecular-weight DNA (Aggarwal *et al.*, 1992).

A quantity of 50 µl of dried blood can provide 0.5 µg DNA, sufficient for multiple PCR-based

assays (McCabe *et al.*, 1987). DNA is stable on cotton cloth for at least 4 years (Gill *et al.*, 1985).

A study planned to verify the stability of dried blood spot specimens for the detection of human immunodeficiency virus DNA by PCR techniques showed that treatment at 37°C and 60% humidity for 7 days, storage for 12 weeks at 22°C, followed by a freeze/thaw cycle had no adverse effect on PCR reactivity when compared with reference spots stored at –20° C. These findings suggest that dried blood spots are a powerful resource for testing for HIV by PCR, especially in remote areas where refrigeration and immediate sample processing are unavailable (Cassol *et al.*, 1992).

Blood components. From 10 ml of blood, the following quantities are obtained:

- plasma (or serum)—6 or 7 ml
- lymphocytes and mononuclear cells—10–20 × 10^6 cells/ml
- erythrocytes (RBCs) and other cells—5 × 10^6 cells/µl; 10–15 mg Hb.

Mononuclear leukocytes are the only cell type in blood capable of growth; they can be cryopreserved for the establishment of cell lines. Cryopreservation permits cell viability (for tissue culture-based mutation assays, transformation for gene mapping, etc.) and can be the only source to measure RNA after PHA induction (RNA is difficult to assay in quiescent cells because levels of activity are absent or low). Granulocytes can serve as a source of DNA without sacrificing the lymphocytes. Erythrocytes, stored after washing with physiological saline, can be useful to study adducts of haemoglobin. Plasma/serum can be used to measure microanalytes, diet components, vitamins, xenobiotic exposures and so on. Plasma can be obtained from an anticoagulated blood sample through separation from cell components. It permits the analysis of coagulation factors, while serum permits better estimate of antibodies, nutrients, and lipid and lipoprotein measurements. Loss of plasma volume because of filtering after fibrin formation is a disadvantage of plasma storage over serum storage.

Pooled aliquots of serum specimens have been used (Wahrendorf *et al.*, 1986) for nutritional or other biochemical studies, e.g. HIV antibody testing (McMahon *et al.*,1995). The approach requires merging aliquots of specimens from persons within a subgroup and testing the combined sample to obtain a group-specific average value. This approach requires only a small number of laboratory tests to be performed, but yields only a mean value without a variance or information about the distribution of results.

DNA can be extracted from whole blood, leukocytes and, as was recently demonstrated (Blomeke *et al.*, 1996), serum, plasma or paraffin-embedded tissues. Small amounts of DNA from blood (sufficient for PCR-based assays) can be obtained through finger-pricks (Guthrie cards), and the subjects can do the collection themselves at home. None the less, some discomfort is often felt with this approach and the participation rate can be reduced. Dried blood spots are very convenient for storage and shipping, and may be optimal for a large-scale epidemiological study. Whole blood obtained directly from finger-pricks may have several applications, e.g. it can be used for analysing complement activation using an ELISA enzyme immunoassay method (Chang and Lister, 1993).

Processing. The sample processing depends on the marker needed. Investigators must design studies to fit the requirements of the critical biomarkers. For catecholamine measurements, for example, blood should be centrifuged within 1 h of collection. Once plasma is prepared, catecholamines are stable for 1 day at 20°C, 2 days at 4°C, 1 month at –20°C, and up to 1 year at –70°C (Boosma *et al.*, 1993). A delay in blood processing may affect the assay results. For example, blood samples held for 2 days at room temperature showed a mean 81% decrease in cells positive for cytomegalovirus antigenaemia (Landry *et al.*, 1995); on the contrary, a delay of up to 24 hours in blood processing did not significantly change the specific activities of arylsulfatase A and cerebroside-beta-galactosidase in both leukocytes and lymphocytes (Shah *et al.*, 1995). The stability of assays in relation to time and temperature of storage has not always been well documented, but should be considered in the context of specific studies. In one study, it was recommended that serum fatty acids be measured within 2 weeks at 4°C, within a few months at –20°C and within 1 year at –80°C to estimate the composition of the major fatty acids (Umemura *et al.*, 1991). Blood samples treated with Triton X-100,

EDTA and sodium fluoride may be stored up to 4 weeks without appreciable effect on measured blood lactate concentration (Hill, 1995)

Storage. It is critical to maintain careful records of the identity and location of all materials, with particular attention to storage history, occurrence of temperature fluctuation and monitoring of stored control specimens, in order to check the effects of storage duration. For example, samples stored on top of a mechanical freezer may be exposed to more extreme temperature fluctuations then those stored at the bottom. This can lead to deterioration of sample over time and can bias the analyses.

In prospective studies based on storage of biological samples at low temperature, a crucial question is to evaluate whether long-term preservation of samples is able to affect the categorization of the subjects involved. A recent study showed that estradiol, free and total testosterone, and prolactin in serum and plasma samples maintained almost the same rank by hormonal concentration throughout a 3-year period of cryoconservation at –80°C (Bolelli *et al.*, 1995). In another study, aliquots from 40 ml plasma pools preserved with metaphosphoric acid were assayed for their ascorbic acid values after 12, 24 and 42 months of storage at –70°C. Similarly, aliquots from 16 plasma pools were assayed for values of retinol, several carotenoids and two tocopherols for a period of storage at –70°C up to 4.3 years. No indications were found of important losses of these antioxidant micronutrients during storage (Comstock *et al.*, 1995)

However, in a study on maternal screening for three markers, immediate freezing of serum and subsequent thawing resulted in a significant increase in beta-hCG and unconiugated E3 levels, but no change in AFP levels. AFP levels were influenced by centrifugation status, and all three analytes were influenced by refrigeration status (Lantz *et al.*, 1995).

Timing. The timing of sample acquisition can profoundly influence the interpretation of results. This is especially evident for those hormones which have hourly, daily and/or monthly cycles. Both timing of exposure and timing of biospecimen collection need to be considered; i.e. mutations cannot be compared with chromosomal aberrations on the same sample, since, for example, chromosomal changes may require a longer time to appear following exposure than DNA-base changes.

It is critical to obtain information (in addition to the usual questionnaire typically administered) at the time of biospecimen collection to aid interpretation. Typical information will be assay-specific, but is likely to include, as a minimum, time and date of draw, volumes and type of specimen, medical illness (current/remote), last food consumed (type and time since collection), medication use, reproductive information—i.e. time since last menstrual period—time since last cigarette, and alcohol intake. Information will differ depending on the study setting and the target population. Some studies may require that specimens be collected from the same persons at several time points; collection and related data should be stored with the study database.

Specific biomarkers will each have a precise time dependence on exposure; examples include sister chromatid exchange (days), chromosome aberrations (around 6 weeks post-exposure), adducts [varies by type, e.g. DNA and albumin: 25 days; Hb: 120 days—see *Environmental Health Perspectives*, Vol. 103 (Suppl. 3), 1995], and various mutations (6 weeks to 6 months after exposure, but also dependent on repair activity). Similar considerations apply to nutrients and hormones, but are less relevant to repeated or remote exposures.

Urine

Because urine is an ultrafiltrate of the plasma, it can be used to evaluate and monitor body homeostasis, metabolic disease processes, exposure to xenobiotic agents, mutagenicity, exfoliated cells, DNA adducts, etc. Urine specimens are usually readily obtainable, although sometimes its collection is felt to be more inconvenient than blood collection. The type of specimen selected and the collection procedure used depend on the tests to be performed. There are basically four types of urine specimens to be collected for epidemiological studies: first morning, random, fractional and timed. The ideal specimen is adequately concentrated to ensure the detection of analyses of interest (for a review, see Brunzel, 1994):

- To collect a first morning specimen, the subject voids before going to sleep and, immediately upon rising, collects a urine specimen. The specimen must be preserved if not delivered

within 2 h of collection. First morning specimens are ideal for substances that require concentration, and white cells, red cells and casts are more stable in concentrated acidic urine specimens. However, they are inadequate for cytology studies; in fact, the epithelial cells may undergo degeneration, and salt crystallization during processing may have adverse effects on cytology studies.

- Random urine specimens can be collected at any time. These specimens are usually satisfactory for routine screening and for cytology studies. One method to increase the cellularity of the urine specimen is to have the subject drink a lot of water 2 h before collection and exercise for 5 min by skipping or jumping up and down prior to specimen collection.
- Fractional collections are used to compare the concentration of an analyte in urine with its concentration in the blood. These specimens are also termed 'double-voided specimens', because the first morning urine (containing solutes and metabolites from the evening meal) is discarded, but the second urine excreted (fasting urine specimen) is collected.
- Timed collections, usually done over 12–24 h periods (or even longer), eliminate the need to determine when excretion is optimal and allow day-by-day comparison of excretion patterns. On day 1, at the start time, the subject empties his/her bladder. For the next 24 h, all subsequent urine must be collected in the container. On day 2, at the end time (the same hour as the start time), the subject empties the bladder and includes this specimen in the collection. Only one first morning sample must be included. Urine collections should be kept on ice or refrigerated throughout the duration of the collection.

Plastic or glass containers must be clean and dry, and have a 50–3000 ml capacity, a wide mouth and a leak-proof screw cap. Depending on the analyte to be measured, a preservative may be needed. The type of preservative may differ according to test methodologies, time delay and transport conditions. In the laboratory, total volume must be recorded, the specimen well mixed to ensure homogeneity, and aliquots removed for the appropriate test.

In epidemiological studies, the period of urine collection used to measure excretion varies depending on the marker to be measured and the aim of the study. For example, using overnight collection to measure catecholamines may result in a much larger sample size requirement to achieve the same power as a 24-h collection. However, 24-h collections incorporate measurement errors that are not present in overnight collections (White *et al.*, 1995). In reference to optimal storage conditions, a study by Boosma *et al.* (1993) showed that catecholamines are stable at 4°C for 1 month in unpreserved urine, and for 4 months in urine preserved with EDTA and sodium metabisulfite. In acidified urine, catecholamines were nearly unchanged after 1 year at 4 and –20°C.

Urine processing can influence the results of various tests. For example, freezing at –20°C and storing of urine samples prior to assessment of albumin concentration can affect the absolute values obtained (Shield *et al.*, 1995). For the quantitation of glutathione transferases by an ELISA procedure, storage of urine samples requires the presence of low detergent concentrations, such as Tween 20, which both stabilizes the enzymes and prevents their absorption by plastic test tubes (Sundberg *et al.*, 1995).

The use of urine requires decisions to be made on the appropriate time frame for measurement of numerous analytes, microbial contamination, the cost of storing large volumes of material and the paucity of studies on the effect of long-term storage on qualitative or quantitative analyte detection.

Tissues

Tumour pathology samples are required to confirm clinical diagnoses by histological analysis. In recent years, there has been increasing interest in examining tumour characteristics at the chromosomal and molecular levels. This generally requires collecting more material than is necessary for pathological evaluation. When possible, the tissue samples should contain tumour as well as normal tissue to permit investigation of the different characteristics of the two tissues. For tissue blocks, collections of stained and unstained tissue samples from the same individual are ideal.

In planning any human tissue banking effort, one should consider which tissues might already be, or will potentially be, available, such as dis-

carded tissue blocks or diagnostic specimens from surgery (informed consent may be required). Some of these tissues may be less than ideal for specific ends, e.g. formalin-preserved samples have altered chemical constituents and PCR assays are difficult. However, with development of new analytical techniques, new uses may be found for any tissues that might be available. Several extraction techniques have been proposed, e.g. the sonication method, which takes only 30 min from start to finish to extract DNA from fresh, frozen and formalin-fixed, paraffin-embedded tissue specimens (Heller *et al.*, 1992). Although available specimens are often limited to formalin-fixed, paraffin-embedded blocks, frozen tissues have distinct advantages. Snap freezing of tissues is especially suited for RNA extraction. Recently, a new method has been developed for preparing normal or tumour tissue for RNA recovery; the tissue sample is placed in a transparent bag which does not break when submerged in liquid nitrogen. While frozen, the tissue is crushed with a hammer. After the specimen is completely lysed, RNA can be extracted even from very limited tissue samples (Gramza *et al.*, 1995)

As an example of the influence of storage temperature on a specific assay, gastric mucosal biopsies stored at 4°C for 1 and 2 weeks resulted in the recovery of 81% and 19% of *Helicobactor pylori*, respectively. Storage at –20°C improved yields to 100% and 57% after 4 and 12 weeks, respectively. Recovery improved still further at –70° C (Han *et al.*, 1995). In general, it is best to use the lowest storage temperature practical, given cost and sample size constraints.

A novel method has been developed for processing biopsy specimens for histochemical and immunohistochemical analyses by combining freezing with low-temperature plastic embedding. The method avoids the need for tissue fixation and combines the superior morphological preservation of fixed embedded tissue with the reactivity of cryostat sections. Biopsy specimens are stored at room temperature without loss of tissue-specific characteristics during storage (Murray & Ewen, 1991).

Adipose tissue

Adipose tissue aspirations may be quite feasible for the subject and involve low risk. Tissue samples can be analysed for assessment of prior exposures in epidemiological studies. This tissue offers a relatively stable deposit of triglyceride and fat-soluble substances, such as fat-soluble vitamins (i.e. vitamins A and D) and pesticides. Halogenated hydrocarbons may be measured in concentrations hundreds of times greater than those in blood of the same individuals. As a tissue, it represents the greatest reservoir of carotenoids in the body, and the tissue composition also reflects long-term dietary intake of essential fatty acids (for a review, see Kohlmeier & Kohlmeier, 1995). Samples are collected from the upper outer quadrant of the buttock while the subject is lying face down. No local anaesthesia is needed. After disinfection of the skin with alcohol, the subject is asked to tense the buttock to delineate muscle and fat. The upper outer quadrant is grasped between the finger and the thumb of one hand. A Luer-lock needle is inserted at a 45° angle, after which a vacuum tube is connected to the needle. The needle is then moved in and out of the adipose tissue, at which time some fat collects at the top of the Luer adapter between the needle and the tube. The adapter is capped and immediately frozen at –80°C (Kohlmeier & Kohlmeier, 1995).

Bronchoalveolar lavage (BAL)

Bronchoalveolar lavage fluid has been used for many years in occupationally related epidemiological studies. A typical example is the use of BAL to assess and quantify asbestos exposure through identification of asbestos bodies and cell content in the alveolar fluid (Orlowski *et al.*, 1994).

Induced sputum samples and bronchoalveolar lavage fluid (BALF) can also provide sufficient DNA for PCR-based assays. As an example, a study by Liesnard *et al.* (1994) showed that PCR can be applied to BALF as a rapid method to detect cytomegalovirus infection.

Bronchoalveolar lavage is performed through a fibre-optic bronchoscope inserted transnasally after topical application of local anaesthesia. Sterile solution (200–250 ml) at 37°C is injected by 50 ml bolus into a dependent subsegmental bronchus of the right middle lobe or lingula. The bronchoalveolar fluid is then recovered by mild aspiration and can be ready for centrifugation, cell count, cell digestion, preparation of slides for asbestos bodies count through light or scanning electron microscopic analyses, etc. (Bell *et al.*, 1981; Roggli *et al.*, 1994).

Exhaled air

Collection and analysis of exhaled air has been done for years to evaluate exposure to different substances, particularly solvents such as benzene, styrene and tetrachloroethylene. In addition, exhaled air samples have been used as a source of exposure and susceptibility biomarkers. An example of this is provided by the [3-13C-methyl] caffeine breath test, developed to measure the P4501A2 activity, which can increase after exposure to polybrominated biphenyls and dioxins (Lambert *et al.*, 1990). After ingestion of a labelled dose of 3 mg/kg caffeine, aliquots of end-tidal breath samples are collected at different times and placed in vacationers for storage and transport. Analysis by differential mass spectroscopy of the samples gives an indirect measurement of the P4501A2 activity. Other examples of tests that exploit labelled compounds with detection of the target compound in expired air include breath methane and H_2 (fermentation), and breath urea (presence of urease positive organisms such as *H. pylori*). For collection of exhaled breath, various similar methods of sampling have been developed; for example, a method capable of measuring sub-ppb levels of volatile organic compounds, employing Tenax sorbents to collect breath samples from a Tedlar bag that has been filled by the subject exhaling through a two-way mouthpiece. The subject inhales pure humidified and charcoal scrubbed air from a cylinder (Wallace & Pellizzari, 1995).

Hair

Hair is an easily available biological tissue whose typical morphology may reflect disease conditions within the body. It provides a permanent record of trace elements associated with normal and abnormal metabolism, as well as those assimilated from the environment (Srivastava & Gupta, 1994). In fact, human hair analysis has proved to be a well-suited biological marker of occupational and environmental exposure to toxic metals (although, because of external contamination, they are not useful for most nutrients). The method of hair analysis can be suitable for use in pilot prospective studies. If an excessive exposure is detected, analysis of conventional biological substrates, such as blood or urine, is recommended in order to verify the exposure accurately (Bencko, 1995).

Hair samples have been proven to be valid biological samples for nicotine measurement in estimating average environmental tobacco smoke (ETS) exposure in children. The hair nicotine levels were shown to be well correlated with cotinine creatinine ratios in urine from the same individuals (Nafstad *et al.*, 1995).

Moreover, hair and urine analyses are complementary tests for establishing drug use. Hair analysis provides long-term information, from months to years, concerning both the severity and pattern of drug use. In contrast, urinalysis only indicates drug use that has occurred within the last 2–3 days. External contamination of hair by drugs present in the environment (e.g. smoke) is the main problem of hair analysis. The problem, however, can be effectively avoided by washing the hair specimen, by kinetic analyses of the wash data and by metabolite measurement. The possibility of bias due to race and/or hair colour is avoided by the exclusion of melanin from hair analysis (Du Pont & Baumgartner, 1995)

Finally, hair roots can be an optimal source of DNA for PCR analysis and permit easy collection, transport and storage and low overall costs (Thompson *et al.*, 1992).

Nail clippings

Toenail or fingernail clippings are obtained in a very easy and comfortable way, and they do not require elaborate processing, storage and shipping conditions and are thus suitable for large epidemiological studies (Garland *et al.*, 1995). Toenail clippings have been used for studies of trace elements and for measurement of selenium levels (which reflect selenium intake) in order to investigate the association between selenium status and cancer risk (Garland *et al.*, 1993; van den Brandt *et al.*, 1993). Another study showed that arsenic levels in fingernails can be a biological indicator of exposure to arsenic (Agahain *et al.*, 1990)

Nail clippings may be less likely to be contaminated by environmental factors, handling procedures or cosmetics (Hambidge, 1982), but involve more complicated washing, specimen hydrolysis and matrix problems for the chemist (Winn *et al.*, 1990).

Buccal cells

As an alternative to using blood obtained by venepuncture, cells may be obtained by an oral

rinse. Human cheek cells were first used as a non-invasive method for detecting tissue lipid profiles in nutritional studies (McMurchie *et al.*, 1984). More recently, a 'swish and spit' technique to collect nucleated cells as a source for DNA has been reported. For this technique, the subject vigorously swishes isotonic saline in the mouth and expectorates it into a collection container. DNA is extracted from the buccal cells and can provide excellent templates for PCR-based assays. In this study, the integrity of the specimens was unaffected by storage at –20, 4, 25 or 37°C. Thus, this process may be extremely useful for large-scale studies that require DNA from subjects geographically distant from the research site (Hayney *et al.*, 1995).

The oral rinse method is perhaps the most extensively used non-blood-based sampling technique. However, it involves liquid sample handling and requires an additional centrifugation step to spin down the cells, increasing the time for DNA preparation. Buccal cells may also be collected on cytology brushes and swabs for use in PCR-based assays. For cytology brushes, there is a variety of collection approaches. Generally, just before specimen collection, the subject should rinse the mouth with water to reduce bacteria interference. We have successfully used a technique involving 30 s of gentle rubbing over the inner buccal surface on each side of the mouth. This method would not be appropriate in subjects with mucosal lesions. Alternatively, a sterile metal spatula can be used to scrape the buccal mucosa firmly but carefully so as not to cause pain or bleeding; the buccal material on the spatula is then spread across the surface of a clean microscope slide. The prepared slide is placed in fixative for approximately 1 h and then allowed to dry at room temperature. Transport may be done in an appropriate box at room temperature. Cheek cell samples and DNA prepared from the swabs are highly stable.

In a blind study comparing the analysis of 12 mutations responsible for cystic fibrosis in multiplex products amplified with DNA from both blood and buccal cell samples from 464 individuals, there was a 100% correlation of results for blood and cheek cells collected on cytology brushes for use in genetic testing (Richards *et al.*, 1993). In this study, the stability of cheek cells was evaluated by collecting cells on both cytology brushes and female Dacron swabs, and storing them at 4°C for 1 month, 2 weeks, 1 week and 3 days. DNA was subsequently prepared from these samples and amplified alongside freshly prepared buccal cell DNA. There was no significant difference in the yield of amplification products among the different samples. To determine how stable the cheek cells would be when shipped under various weather conditions, samples were stored for 3 days in a –20°C freezer, a 37°C incubator and in the airspace of a 37°C water bath, in order to simulate the effect on buccal cell stability of very cold, hot or humid conditions. Amplified product yields were equivalent to those of freshly collected and extracted samples.

Buccal mucosal transudate collected by salt-impregnated buccal swab was shown to provide a minimally uncomfortable measure of cocaine use in a preliminary study of 44 subjects (Leonard *et al.*, 1994). Buccal cells can be used to count micronuclei. The proportion of exfoliated buccal cells with micronuclei offers the opportunity to assess sensitivity to gamma radiation and genotoxic compounds and, in addition, to monitor the effectiveness of cancer intervention strategies (Belien *et al.*, 1995).

Saliva

Saliva can be an efficient, painless and relatively inexpensive source of biological materials for certain assays, and provides a useful tool for measuring endogenous and xenobiotic compounds. Several devices, such as non-covered cotton roll, polypropylene-covered polyether roll and paraffin wax chewing stimulation, have been developed to collect effective saliva samples. The different devices, which have been used to measure corticosteroids (Kathol *et al.*, 1995), antibodies to human immunodeficiency virus type 1 (HIV-1) (Granade *et al.*, 1995), and other compounds, generate different data. Care should be taken when different devices are employed (Lenander-Lumikari *et al.*, 1995). Recently, a technique has been described to determine the CYP1A2 phenotype following a caffeine dose using saliva (Fuhr & Rost, 1994).

Cotinine determination in saliva is considered a reliable marker for both environmental tobacco smoke exposure and active smoking (Samet *et al.*, 1988).

Breast milk

Several studies have been conducted using breast milk from lactating mothers to measure hormones, epoxides of cholesterol (Wrensch *et al.*, 1993), exposure to chemicals, i.e. polychlorinated biphenyls (Becker *et al.*, 1995, Schlaud *et al.*, 1995), pesticides (Rogan & Ragan, 1994) and biological contaminants, i.e. aflatoxin M1 (El-Nezami *et al.*, 1995) or *Borrelia burgdorferi* (Schmidt *et al.*, 1995).

In studies of the association between selenium levels and cancer risk (Sanz & Diaz, 1995), breast milk has been used to estimate serum selenium levels, because it is well correlated to selenium intake. Milk samples are collected when breast feeding is established. The samples are obtained by expression, either manually or by vacuum pump. Before expressing the milk, the subject should wash the nipples, and then the milk should be expressed directly into autoclaved glass bottles, which are opened and used for one expression only. After collection, milk can be kept in a domestic refrigerator until it is sent to the laboratory (within 24 h), where it can be deep frozen for long-term storage.

Faeces

Certain cells of interest (Nair *et al.*, 1994), infectious markers (Ramamurthy *et al.*, 1993), oncogenes (Celani *et al.*, 1993; Mao *et al.*, 1994), RNA (Davidson *et al.*, 1995) and certain specific compounds may best be studied in faeces, e.g. faecapentaenes, which are potent mutagens excreted in faeces. In one study, a 2-day stool specimen was collected using a protocol that required subjects to place dry ice into a plastic container held by a collection bonnet positioned on the toilet rim. The container was then placed into a styrofoam dry ice chest. Subjects were instructed on how to avoid contamination with urine. Frozen stools underwent lyophilization prior to analysis. The procedure was successful in that neither occult bleeding, laboratory drift nor sample degradation with storage (re-test after 2 years) influenced the assay. The correlation between first and replicate measures was greater than 0.90 (Schiffman *et al.*, 1989).

Semen

Semen specimens are collected to evaluate the effects of exposures on endocrine and reproductive factors. Sexual abstinence for at least 2 days, but not exceeding 7, should precede the collection. The specimen is collected by subjects through masturbation, and the entire ejaculate is collected in a clean sterile plastic or glass container. Lubricants and ordinary condoms should not be used because of their spermicidal properties. The collection container should be at room temperature or warmed before collection, should reach the laboratory within 1 h of collection, should be protected from extreme temperatures (i.e. maintained between 20 and 40°C), and should be labelled with the time of actual collection, a crucial factor in evaluating liquefaction and sperm motility (World Health Organization, 1992).

Temperature

Specimen collection requires storage systems that are capable of maintaining the optimal temperature for the diverse types of specimens:

- –20°C: certain items stable, i.e. urine
- –70°C:
  - —cell viability limited (not optimal)
  - —DNA stable
  - —serum stable
  - —most hormones stable
  - —most vitamins stable
- –120°C: hormones, carotenoids, other nutrients.

Liquid nitrogen, present on the bottom of the freezer, is at –196°C, while the area higher up in the freezer, i.e. at the liquid/vapour interface is typically around –120°C. If samples are stored at both temperatures, analyses must be stratified and different storage procedures taken into account.

Most investigators use liquid nitrogen to maintain cell viability. However liquid nitrogen can be expensive and labour-intensive on a large scale. In recent years, –140°C freezers with liquid nitrogen back-up systems have been developed as an alternative to storage in liquid nitrogen. The scientific rationale for using these freezers as an alternative, is that, below –139°C for pure water (or about –130°C for culture media), molecules still vibrate, but do not move from one position to another, thus preventing chemical reactions. Therefore, few, if any, changes that may affect cell viability should occur between –140 and –196°C.

Freezers may fail, leading to the necessity for 24-h monitoring of the facility through a computerized alarm system to alert personnel and activate back-up equipment. For large-scale collections, systems

must be in place that monitor for the possibility of fire, power loss, leakages, breaches of security, etc. An empty functioning freezer than can accept samples after a single failure is crucial. Monitoring should include active 'hands-on' manual and mechanical checks, since monitoring systems themselves are prone to failure (e.g. a temperature gauge can become 'stuck' in place). Receipt and control procedures, a quality assurance programme including plans for routine preventive maintenance, and a system for documenting the storage history of every specimen must be carefully planned.

Shipping

Sample shipping requirements depend on the time, distance, climate, season, method of transport, applicable regulations, type of specimen and marker(s) to be assessed. Usually, polyurethane boxes containing dry ice are used to ship and transport samples that require low temperatures. The quantity of dry ice should be carefully calculated, based on the estimated time of the trip (dry ice evaporates with time, depending on the external temperature), the number of samples to be transported in the boxes and an ample safety factor that takes into account likely delays. If the boxes have to be shipped by airline, it is suggested that they be placed in the hold of the plane, where the temperature during the trip is very low. Each box should be accompanied by a typed chart describing in detail the contents of the box and the location of each tube in the box.

For samples that require very low temperatures, shipping in a liquid nitrogen container can be optimal. If international shippings are planned, safety declarations for the foreign country's customs should be prepared.

Safety

No discussion of biospecimens is complete without a mention of biosafety concerns, although a complete discussion goes beyond the scope of this brief review. Historically, the earliest concern was to protect the biospecimen from contamination; however, in recent years, the appreciation of potentially deadly blood-borne pathogens such as HIV and hepatitis has radically re-oriented this approach towards worker safety. Changes to laboratory design (e.g. air flow is 'once through' and not recirculated, laminar flow hoods, restricted entrance), equipment (easy cleaning benchtops, easily available handwashing and first aid such as eye washing fountains) and techniques (forbidding mouth pipetting, puncture-resistant containers for disposal of sharps, mandatory training, use of laboratory coats, etc.) have been made. Inoculation of all workers who may be involved with human biospecimens with hepatitis B vaccine and 'universal precautions' training should be standard.

Procedures

It is appropriate to conclude this survey by emphasizing that epidemiological studies require standardized approaches in order to ensure quality control. All epidemiologists are familiar with edit checks on data entry and similar measures used to ensure the integrity of data. A similar approach with regard to dealing with biospecimens is essential. For example, our 'biospecimen collection manual' contains numbered steps (each with a rationale) for accomplishing the routine tasks required for pre-clinic, clinic and post-clinic procedures involved with biospecimen acquisition. Field trip preparation, packing and shipping samples, etc. are dealt with in a similar manner. Procedures are in place to standardize approaches at laboratories that perform routine tasks such as DNA extraction and at repositories that receive, record, process, store, label and ship samples. Flow charts are often included to highlight decision points, i.e. 'DNA to lab A', 'serum to lab B'. These procedures are invaluable in guiding day-to-day operations, assisting with the training of new personnel, serving as a starting point for the evaluation of technical improvements, creating a framework for periodic reviews and establishing a baseline for the evaluation of a response to untoward events.

References

Agahain, A., Lee, J.S., Nelson, J.H. & Johns, R.E. (1990) Arsenic levels in fingernails as biological indicator of exposure to arsenic. *Am. Ind. Hyg. Assoc. J.*, 51, 646–651

Aggarwal, R.K., Lang, J.W. & Singh, L. (1992) Isolation of high-molecular-weight DNA from small samples of blood having nucleated erythrocytes, collected, transported, and stored at room temperature. *Genet. Anal. Tech. Appl.*, 9, 54–57

Becher, G., Skaare, J.U., Polder, A., Sletten, B., Rossland,

O.I., Hansen, H.K. & Ptashekas, J. (1995) PCDDs, PCDFs, and PCBs in human milk from different parts of Norway and Lithuania. *J. Toxicol. Environ. Health.*, 46, 133–148

Belien, J.A., Copper, M.P., Braakhuis, B.J., Snow, G.B. & Baak, J.P. (1995) Standardization of counting micronuclei; definition of a protocol to measure genotoxic damage in human exfoliated cells. *Carcinogenesis*, 16, 2395–2400

Bell, D.Y., Haseman, J.A., Spock, A., McLennan, G. & Hook, G.E.R. (1981) Plasma proteins of the bronchoalveolar surface of the lungs of smokers and nonsmokers. *Am. Rev. Respir. Dis.*, 124, 72–79

Bencko, V. (1995) Use of human hair as a biomarker in the assessment of exposure to pollutants in occupational and environmental settings. *Toxicology*, 101, 29–39

Bernard, A.M. (1995) Biokinetics and stability aspects of biomarkers: recommendations for application in population studies. *Toxicology*, 101, 65–71

Blomeke, B., Bennett, W.P., Harris, C.C. & Shields, P.G. (1996) Serum, plasma, and paraffin embedded tisssues as sources of DNAs for studying genetic susceptibility to cancer. *Proc. Am. Assoc. Canncer Res.*, 37, 261

Bolelli, G., Muti, P., Micheli, A., Sciajno, R., Franceschetti, F., Krogh, V., Pisani, P. & Berrino, F. (1995) Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma. *Cancer Epidemiol. Biomarkers Prev.*, 4, 509–513

Boosma, F., Alberts, G., Van-Eijik, L., Man-in-'t-Veld, A.J. & Schalekamp, M.A. (1993) Optimal collection and storage conditions for catecholamine measurements in human plasma and urine. *Clin. Chem.*, 39, 2503–2508

Bruzel, N. (1994) *Fundamentals of Urine and Body Fluids Analysis*, Philadelphia, PA, WB Saunders

Cassol, S., Salas, T., Gill, M.J., Montpetit, M., Rudnik, J., Sy, C.T. & O'Shaugnessy, M.V. (1992) Stability of dried blood spot specimens for detection of human immunodeficiency virus DNA by polymerase chain reaction. *J. Clin. Microbiol.*, 30, 3039–3042

Celani, P., Berchtold, C.M., Mahbry, M., Carroll, M., Sidransky, D., Cesano, R.A., Jr & Lupo, R. (1993) Induction of markers of normal differentiation in human colon carcinoma cells by the v-ras-H oncogene. *Cell Growth Diff.* 4, 341–347

Chang, T.M. & Lister, C.W. (1993) Use of finger-prick human blood samples as a more convenient way for invitro screening of modified hemoglobin blood substitutes for complement activation: a preliminary report. *Biomater. Artif. Cells Immobilisation Biotechnol.*, 21, 685–690

Comstock, G.W., Norkus, E.P., Hoffman, S.C., Xu, M.-W. & Helzsouer, K.J. (1995) Stability of ascorbic acid, carotenoids, retinol, and tocopherols in plasma stored at –70°C for 4 years. *Cancer Epidemiol. Biomarkers Prev.*, 4, 505–507

Davidson, L.A., Jiang, Y.H., Lupton, J.R. & Chapkin, R.S. (1995) Noninvasive detection of putative biomarkers for colon carcinoma using fecal messenger RNA. *Cancer Epidemiol. Biomarkers Prev.*, 4, 643–647

DuPont, R.L. & Baumgartner, W.A. (1995) Drug testing by urine and hair analysis: complementary features and scientific issues. *Forensics Sci. Int.* 70, 63–76

El-Nezami, H.S., Nicoletti, G., Neal, G.E., Donohue, D.C. & Ahokas, J.T. (1995) Aflatoxin M1 in human breast milk samples from Victoria, Australia and Thailand. *Food Chem. Toxicol.*, 33, 173–179

Fuhr, U. & Rost K.L. (1994) Simple and reliable CYP1A2 phenotyping by the paraxanthine/caffeine ratio in plasma and in saliva. *Pharmacogenetics*, 4, 109–116

Garland, M., Morris, J.S., Rosner, B.A., Stampfer, M.J., Spate, V.L., Baskett, C.J., Willett, W.C. & Hunter, D.J. (1993) Toenail trace element levels as biomarkers: reproducibility over a 6-year period. *Cancer Epidemiol. Biomarkers Prev.*, 2, 493–497

Garland, M., Morris, J.S., Stampfer, M.J., Colditz, G.A., Spate, V.L., Baskett, C.K., Rosner, B., Speizer, F.E., Willett, W.C. & Hunter, D.J. (1995) Prospective study of toenail selenium levels and cancer among women. *J. Natl Cancer Inst.*, 87, 497–505

Gill, P., Jeffrey, A.J. & Werrett, D.J. (1985) Forensic application of DNA "fingerprints". *Nature*, 318, 577–579

Gramza, A.W., Lucas, J.M., Mountain, R.E., Schuller, D.E. & Lang, J.C. (1995) Efficient method for preparing normal and tumor tissue for RNA extraction. *BioTecniques*, 18, 227–228

Granade, T.C., Phillips, S.K., Parekh, B., Pau, C.P. & Georgge, J.R. (1995) Oral fluid as a specimen for detection and confirmation of antibodies to human immunodeficiency virus type 1. *AADE Ed. J.*, 2, 395–399

Hambidge, K.M. (1982) Hair analysis: worthless for vitamins, limited for minerals. *Am. J. Clin. Nutr.*, 36, 943–949

Han, S.W., Flamm, R., Hchem, C.Y., Kim, H.Y., Clarridge, J.E., Evans, D.G., Beyer, J., Drnec, J. & Graham, D.Y. (1995) Transport of storage of *Helicobacter pylori* from gastric mucosal biopsies and clinical isolates. *Eur J. Clin. Microbiol. Infect. Dis.*, 14, 349–352

Hayney, M.S., Dimanling, P., Lipsky, J.J. & Poland, G.A. (1995) Utility of a "swish and spit" tecnique for the collection of buccal cells for TAP haplotype determination. *Mayo Clin. Proc.*, 70, 951–954

Hill D.W. (1995) Effect of storage on measured blood lac-

tate concentration: a research note. *Int. J. Sports Med.*, 16, 88–90

Heller, M.J., Robinson, R.A., Burgart, L.J., TenEyck, C.J. & Wilke, W.W. (1992) DNA extraction by sonication: a comparison of fresh, frozen, and paraffin-embedded tissues extracted for use in polymerase chain reaction assays. *Mod. Pathol.*, 5, 203–206

Kaaks, R., van der Twel, I., van Noord, P.A. & Riboli, E. (1994) Efficient use of biological banks for biochemical epidemiology: exploratory hypothesis testing by means of a sequential t-test. *Epidemiology*, 5, 429–438

Kathol, R.G., Poland, R.E., Stokes, P.E. & Wade, S. (1995) Relationship of 24-hour free cortisol to 4-hour salivary morning and afternoon cortisol and cortisone as measured by a time-integrated oral diffusion sink. *J. Endcrinol. Invest.*, 18, 374-377

Kieldsen, S.E., Eide, I. & Leren, P. (1983) Effects of posture on serum cholesterol fractions, cholesterol ratio and triglycerides. *Scan. J. Clin. Lab. Invest.*, 43, 119–121

Kohlmeier, L. & Kohlmeier, M. (1995) Adipose tissue as a medium for epidemiologic exposure assessment. *Environ. Health Perspect.*, 103 (suppl. 3P), 99–106

Lambert, G.H., Schoeller, D., Humphrey, H.E., Kotake, A.N., Lietz, H., Campbell, M., Kalow, W. & Spielberg, S. (1990) The caffeine breath test and caffeine urinary metabolite ratio in the Michigan cohort exposed to polybrominated biphenyls: a preliminary study. *Environ. Health Perspect.*, 89, 175–181

Landry, M.L., Ferguson, D., Cohen, S., Huber, K. & Wetherill, P. (1995) Effect of delayed specimen processing on cytomegalovirus antigenemia test results. *J. Clin. Microbiol.*, 33, 257–259

Lantz, M.E., O'Brien, W.F. & William, M.C. (1995) The effect of sample preparation and storage on maternal triple-marker screening. *Obstet. Gynecol.*, 85, 919–923

Lenander-Lumikari, M., Johansson, I., Vilja, P. & Samaranayake, L.P. (1995) Newer saliva collection methods and saliva composition: a study of two salivette kits. *AADE Ed. J.*, 1, 86–91

Leonard, J., Doot, M., Martin, C. & Airth-Kindree, W. (1994) Correlation of buccal mucosal transudate collected with a buccal swab and urine levels of cocaine. *J. Addict. Dis.*, 13, 27–33

Liesnard, C., De Wit, L., Motte, S., Brancart, F. & Content, J. (1994) Rapid diagnosis of cytomegalovirus lung infection by DNA amplification in bronchoalveolar lavages. *Mol. Cell Probes*, 8, 273–283

McCabe, E.R.B., Huang S.-Z, Seltzer, W.K. & Law, M.L. (1987) DNA microextraction from dried blood spots on filter paper blotters: potential applications to newborn screening. *Hum. Genet.*, 75, 213–216

McMahon, K.J., Fang, C., Layug, L. & Sandler, S.G. (1995) Pooling blood donor samples to reduce the cost of HIV-1 antibody testing. *Vox Sang.*, 68, 215–219

McMurchie, E.J., Potter, J.D., Rohan, T.E. & Hetzel, B.S. (1984) Human cheek cells: a non-invasive method for determininfg tissue lipid profiles in dietary and nutritional studies. *Nutr. Rep. Int.*, 29, 519–526

Mao, L., Hruban, R.H., Boyle, J.O., Tockman, M. & Sidransky, D. (1994) Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. *Cancer Res.*, 54, 1634–1637

Murray, G.I. & Ewen, S.W. (1991) A novel method for optimum biopsy specimen preservation for histochemical and immunohistochemical analysis. *Am. J. Clin. Pathol.*, 95, 131–136

Nafstad, P., Botten, G., Hagen, J.A., Zahlsen, K., Nilsen, O.G., Silsand, T. & Kongerud, J. (1995) Comparison of three methods for estimating environmental tobacco smoke among children aged between 12 and 36 months. *Int. J. Epidemiol.*, 24, 88–94

Nair, G.B., Shimada, T., Kurazono, H., Okuda, J., Pal, A., Kavasawa, T., Mihara, T., Uesaka, Y., Shirai, H. & Garg, S. (1994) Characterization of phenotype, serologic, and toxicologic traits of *Vibrio Cholerae* 0139 bengal. *J. Clin. Microbiol.*, 32, 2775–2779

Nielsen, E.W., Johansen, H.T., Straume, B. & Mollnes, T.E. (1994) Effects of time, temperature and additives on a functional assay of C1 inhibitor. *J. Immunol. Methods*, 173, 245–251

Orlowski E. Pairon, J.C., Ameille, J., Janson, X., Iwasubo, Y., Dufour, J., Bignon, J. & Brochard, P. (1994) Pleural plaques, asbestos exposure and asbestos bodies in brochoalveolar fluid. *Am. J. Ind. Med.*, 26, 349–358

Ramamurthy, T., Pal, A., Bag, P.K., Bhattacharya, S.K., Nair, G.B., Kurazono, H., Yamasaki, S., Shirai, H., Takeda, T. & Vesaka, Y. (1993) Detection of cholera toxin gene in stool specimen by polymerase chain reaction: comparison with bead enzyme linked immunosorbent assay and culture method for laboratory diagnosis of cholera. *J. Clin. Microbiol*, 31, 3068–3070

Richards, B., Skoletsky, J., Shuber, A.P., Balfour, R., Stern, R.C., Dorkin, H.L., Parad, R.B., Witt, D. & Klinger, K.W. (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Human Mol. Genetics*, 2, 159–163

Rogan, W.J. & Ragan N.B. (1994) Chemical contaminants, pharmacokinetics, and the lactating mother. *Environ. Health Perspect.*, 102 (suppl. 11P), 89–95

Roggli, V.L., Coin, P.G., MacIntyre, N.R. & Bell D.Y.

(1994) Asbetsos content of bronchoalveolar fluid. A comparison of light and scanning electron microscopic analysis. *Acta Cytologica*, 38, 502–510

Samet, J.M., Marbury, M.C. & Spengler, J.D. (1988) State of art: health effects and sources of indoor air pollution. *Am. Rev. Respir. Dis.*, 137, 221–242

Sanz, A.M. & Diaz, R.C. (1995) Selenium in human lactation. *Nutr. Rev.*, 53, 159–166

Schiffman, M.H., Van Tassell, R.L., Robinson, A., Smith, L., Daniel, J., Hoover, R.N., Weil, R., Rosenthal, J., Nair, P.P., Schwartz, S., Pettigrew, H., Curiale, S., Batist, G., Block, G. & Wilkins, T.D. (1989) Case–control study of colorectal cancer and fecapentaene excretion. *Cancer Res.*, 49, 1322–1326

Schmidt, B.L., Aberer, E., Stockenhuber, C., Klade, H., Breier, F. & Luger, A. (1995) Detection of *Borrelia burgdorferi* DNA polymerase chain reaction in the urine and breast milk of patients with Lyme borreliosis. *Diagn. Microbiol. Infect. Dis.*, 21, 121–128

Shah, S.N., Kohnson, R.C., Minn, K. & Schoenfeld, F. (1995) Arylsulfatase A and beta-galactosidase activities in leukocytes and lymphocytes from normal and psychiatric subjects. Effects of blood processing delay and interleukin-2 stimulation. *Mol. Chem. Neuropathol.*, 24, 43–52

Shield, J.P., Hunt, L.P., Morgan, J.E. & Pennock, C.A. (1995) Are frozen urine for estimating samples acceptable albumin excretion in research? *Diabet. Med.*, 12, 713–716

Srivastava, A.K., Gupta, B.N. (1994) The role of human hairs in health and disease with special reference to environmental exposures *Vet. Hum. Toxicol.*, 36, 556–560

Statland, B.E., Winkel P. & Harris, S.C. (1978) Evaluation of biological sources of variation of leukocytes count and other hematologic quantities using very precise automated analyzers. *Am. J. Clin. Pathol.*, 69, 48–54

Storch, G.A., Gaudreault-Keener, M. & Welby, P.C. (1994) Comparison of heparin and EDTA transport tubes for detection of cytomegalovirus in leukocytes by shell vial assay, pp65 antigenemia assay, and PCR. *J. Clin. Microbiol.*, 32, 2581–2583

Sundberg, A.G., Nilsson, R., Appelkvist, E.L. & Dallner, G. (1995) ELISA procedure for the quantitation of glutathione transferases in the urine. *Kidney Int.*, 48, 570–575

Thompson, D.M., Brown, N.N. & Clague, A.E. (1992) Routine use of hair root or buccal swab specimens for PCR analysis: advantages over using blood. *Clin. Chim. Acta*, 207, 169–174

Umemura, U., Koike, K.A., Sato, S., Iso, H., Shimamoto, T. & Komachi, Y. (1991) The effect of storage on serum fatty acids. *Nippon Eisegaku Zasshi*, 46, 976–983

van den Brandt, P.A., Goldbohm, R.A., van't Veer, P., Bode, P., Hermus, R.J.J. & Sturmans, F. (1993) Predictors of toenail selenium in men and women. *Cancer Epidemiol. Biomarkers Prev.*, 2, 107–112

Wahrendorf, J., Hanck, A.B. & Munoz, N. (1986) Vitamin measurements in pooled blood samples. *Am. J. Epidemiol.*, 123, 544–550

Wallace, L.A. & Pellizzari, E.D. (1995) Recent advances in measuring exhaled breath and estimating exposure and body burden for volatile compounds (VOCs). *Environ. Health Perspect.*, 103 (suppl. 3), 95–98

White, I.R., Brunner, E.J. & Barron, J.L. (1995) A comparison of overnight and 24 hour collection to measure urinary catecholamines. *J. Clin. Epidemiol.*, 48, 263–267

Winn, D.M., Reichman, M.E. & Gunter, E. (1990) Epidemiologic issues in the design and use of biological specimen banks. *Epidemiol. Res.*, 12, 56–70

World Health Organization (1992) *Laboratory Manual for the Examination of Human Semen and Semen-cervical Mucus Interaction* (3rd edn), New York, Cambridge University Press, pp. 5–6

Wrensch, M., Petrakis, N.L., King, K.B. & Lee, M.M. (1993) Breast cancer risk associated with abnormal cytology in nipple aspirates of breast fluid and prior history of breast biopsy. *Am. J. Epidemiol.*, 137, 829–833

Corresponding author

M.T. Landi
Genetic Epidemiology Branch, EPN 400
National Cancer Institute, National Institutes of Health
6130 Executive Blvd., Bethesda, MD 20892-7360,
USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Issues involving biomarkers in the study of the genetics of human cancer

N. Caporaso and A. Goldstein

The investigation of hereditary factors in human cancer was suggested from kindreds that exhibited aggregations of cancer consistent with Mendelian inheritance. A subset of cancer that exhibits strong familial tendencies is due to single genes that 'cause' cancer; more commonly, hereditary factors may influence tumorigenesis in a stepwise probabilistic rather than deterministic manner through a variety of mechanisms, e.g. influencing the disposition of carcinogens. The roles of both common susceptibility genes and rare 'familial' cancer genes are receiving increasing attention in the general population. Population-based studies designed to examine more common genetic variants differ from linkage-based studies. Candidate susceptibility genes may be studied by phenotype or genotype approaches, and the relative advantages and disadvantages of each approach are considered. The issue of gene–environment interaction, implicit in the concept of susceptibility genes, is considered. The influence of genetic factors on individual and attributable risk is addressed.

Cancer and genetics

The general evidence in support of an important role for genes in cancer derives from various interrelated areas: (1) the discovery of specific chromosomal defects associated with lymphoproliferative (and later other) cancers; (2) the universal observation of somatic mutations in tumours; (3) the role of dominant oncogenes and tumour suppresser genes in certain tumours; (4) the mapping and cloning of genes specifically accounting for a number of familial cancers; (5) the general observation of increased tumour incidence in relatives of individuals with cancer compared to suitable controls; (6) the relationship between somatic and hereditary mutations in specific tumours such as retinoblastoma; (7) the increased incidence of tumours associated with conditions that damage the genetic material, e.g. the 'chromosome instability syndromes'; (8) the fact that many established carcinogens are known to damage or disrupt DNA. These lines of evidence can be considered to have clearly established a central role for genetic processes in neoplasia.

Likewise, environmental factors are considered causal for a variety of cancers. A few examples are tobacco and lung cancer, aromatic amine exposure and bladder cancer, and diethylstilbesterone and clear cell carcinoma of the vagina in daughters of exposed women. Our view is that the development of cancer is a complex multistage process, with each step involving a variable mix of environmental and genetic influences. Our knowledge of how these factors act in concert with genes to cause malignancy is currently a patchwork. Three general areas where further work is needed are: (1) etiological factors responsible for certain cancers remain unknown (e.g. prostate, brain); (2) even for cancers where the environmental agents are well established, determinants of individual susceptibility are not clear; (3) the mechanism of interaction between exposures and genes is poorly understood.

Approaches to the study of genetics and cancer in human populations

We first briefly consider approaches that have been used to study highly penetrant 'single' genes associated with familial cancers. We will then describe population-based studies of low penetrant, but more common, 'susceptibility' genes, thought to contribute to complex disorders. Studies of somatic gene findings will be briefly described.

Table 1. Genes associated with familial cancer that have been mapped or cloned

Syndrome (cancer)	Gene	Chromosome	Reference
Retinoblastoma (retinoblastoma, osteosarcoma)	*RB1*	13q14	Cavene *et al.*, 1983; Sparkes *et al.*, 1983
Wilms' tumour (WAGR[a])	*WT2* (*WT1*)	11p13-14	Garwin *et al.*, 1995
Neurofibromatosis type 1 (neurofibrosarcomas, others)	*NF1*	17q11	Upadhya *et al.*, 1995
Neurofibromatosis type 2 (bilateral vestibular schwannomas)	*NF2*	22q11-13	Seizinger *et al.*, 1987; Wertalecki *et al.*, 1988
Li-Fraumeni (breast, sarcoma, leukaemia, brain)	*p53*	17p13.1	Malkin *et al*, 1990
Familial breast–ovary (breast, ovary)	*BRCA1* *BRCA2*	17q21-3 13q12-13	Hall *et al.*, 1990; Albertsen *et al.*, 1994; Miki *et al.*,1994; Wooster *et al.*,1994
Von Hippel–Lindau (renal cell carcinoma, haemangioblastoma)	*VHL*	3p	Seizinger, 1988
CMM (melanoma)	*p16/CDKN2A* *CDK4*	9p21 12q13	Kamb *et al.*, 1994; Hussussion *et al.*, 1994; Zuo *et al.* 1996
NBCC (basal cell, fibrosarcoma, medulloblastoma)	*PTCH*	9q22	Hahn, 1996; Johnson *et al.*, 1996
Ataxia telangiectasia (lymphoma, leukaemia, breast, others)	*ATM*	11q22-q23	Savitsky *et al.*, 1995
Familial adenomatous polyposis/ Gardner syndrome (colon)	*APC*	5q21	Bodmer *et al.*, 1987; Kinzler *et al.*, 1991; Nagase *et al.*, 1992; Eckert *et al.*, 1994
Xeroderma pigmentosa (squamous cell carcinomas, skin, adrenal)	*XP* (A–G) many	9q34.1, others	Tanaka *et al.*, 1990
HNPCC (colon)	*hMSH2* *hMLH1* *hPMS1* *hPMS2*	2p22 3p21 2q31-33 7p22	Nicolaides et al., 1994; Nystrom-Lahti *et al.*, 1994; Papadopolous *et al.*, 1994
Multiple endocrine neoplasia type 1 (carcinoids, pancreas, parathyroid, pituitary)	*MEN1*	11q 13	Larsson, 1988; Thakker *et al.*, 1993
Multiple endocrine neoplasia, type 2A (medullary thyroid, phaeochromocytoma), type 2B (same as 2A)	*RET*	10q11.2	Muligan *et al.*, 1993

[a] WAGR, Wilms' tumour, aniridia, genital, renal abnormalities.

Familial cancer and 'single' gene traits

Cancer families have been noted for centuries and family aggregation studies have generally focused on three related questions. Does a significant excess of cancer exist? Is the cause consistent with genetic or environmental causes, or both? If genetic, what is the mechanism of inheritance?

Recognizing rare syndromes that exhibit strong familial aggregations implicated heritable elements in these entities. An initial step is often to test the transmission pattern of disease in families for a fit with Mendelian patterns of inheritance (segregation analysis). Once a genetic pattern is considered likely, linkage studies or other approaches can be used to map the chromosomal location, to be followed by molecular genetic studies to identify the specific genes involved. For certain familial cancers (e.g. retinoblastoma), as well as various inherited syndromes that often include cancer (e.g. Beckwith-Wiedemann and Wilms' tumour), either the chromosomal location has been identified or the gene has been cloned, or both (Rowley, 1980; Evans, 1993; Knudson, 1993) (Table 1). Gene mapping depends on the availability of suitable multigeneration families or sibships, accurate diagnosis, biospecimens from which DNA may be isolated, and adequate informative markers (generally increasingly available). 'Informative' in this context means that the parent (at least one, depending upon the inheritance pattern) must be heterozygous at both marker and disease loci. DNA markers are characterized in the laboratory and suitable analytical approaches, software and computational resources are applied to evaluate the 'recombination fraction' (evidence of 'crossing over' between a marker and disease locus). The resulting statistical evidence is expressed as a 'lod' score or log of the odds that a given marker is 'linked' to the cancer, with scores of greater than 3 (1000:1) indicative of linkage (Ott, 1991). Once a suspect location has been identified by this approach, positional cloning (also termed 'reverse genetics') can be used to precisely locate and clone the gene. Other approaches are possible. For instance, beginning with a particular protein, the amino acid sequence is used to infer the base pair sequences. Using in-situ hybridization, the chromosomal location may be deduced. Table 1 indicates a number of the major cancer genes identified by these approaches.

Studies involving somatic genes

Chromosome studies

It was recognized in early studies that chromosome aberrations accumulate with the age of a tumour (Boveri, 1917). With the development of better techniques to image chromosomes (i.e. staining of metaphase spreads), it was appreciated that specific chromosome aberrations characterized certain severe multisystem birth defects such as Down syndrome (trisomy 21). In 1960, Nowell & Hungerford (1960) discovered a consistent change in the morphology of a chromosome in blood cells of patients with chronic myelocytic leukaemia (Philadelphia chromosome). We understand today that this translocation involves an exchange of pieces between chromosomes 9 and 22. The break moves the *abl* oncogene from chromosome 9 to the vicinity of the *bcr* (breakpoint cluster region) gene on chromosome 22. When these two genes are brought into juxtaposition, increased tyrosine kinase activity results, and enhanced cell division results in proliferation of cells with this clone. A variety of other translocations involved with specific (typically haematopoietic) tumours involve similar mechanisms.

Chromosome findings have general importance for cancer. The consistent and specific cytogenetic changes observed with particular tumours are an important indication that the genetic material is fundamentally involved in the neoplastic process. Chromosome aberrations in familial cancers can provide clues to the location of regions where critical genes may be located. Although the early techniques for studying chromosomes were difficult and time-consuming, early observations of these consistent defects have led to a more precise understanding of molecular pathology at the DNA level. New approaches, such as fluorescent in-situ hybridization (FISH) technology, are revolutionizing the field and rendering the recognition of syndromes characterized by chromosome findings much simpler.

Other somatic gene findings

Somatic mutations are universally observed in human cancer and are distinguished from hereditary 'single' or 'susceptibility' gene changes discussed previously in that they are observed in tumour tissue (as opposed to germline DNA) and they are not transmitted to offspring.

Loss of heterozygosity studies have been used to identify regions of gene loss to implicate possible tumour suppresser gene loci for many tumours, including chromosome 13 in retinoblastoma (13q14) and chromosome 3 (3p) in lung cancer (Yokota *et al.*, 1987). These studies complement chromosome studies and, together with these, have helped to localize the chromosome abnormalities that characterize various tumours. For example, cytogenetic studies led to the identification of 3p deletions in small cell lung cancer (Peng *et al.*, 1982), an area thought to include more than one tumour suppresser gene, including the recently identified *FHIT* gene (Sozzi *et al.*, 1996).

More recent studies have emphasized the relationship of the type, number and specificity of these and other molecular findings to previous exposure history, hereditary factors, tumour type and aggressiveness. Tumour tissue mutations are characterized by special stains (i.e. immunostaining for *p53*) or direct sequencing of DNA extracted from frozen tissue, or, as is increasingly the case, from tumour blocks. Typically, slides cut from adjacent material have undergone pathological verification. Often, microdissection is used to isolate tumour tissue for special study.

While molecular studies of human tumours have proceeded since the advent of the techniques, the variable nature of many findings has led to efforts to examine hypotheses more rigorously using the methods of epidemiology. Three examples of these approaches are considered.

First, it is hypothesized that somatic mutations may reflect specific exposures that caused the cancer. An example is the relationship between aflatoxin B1 and G to T transversions at codon 249 of *p53* in hepatocellular cancer from certain areas of China or Africa (Soini *et al.*, 1996). Less specifically, and perhaps consistent with the multiple carcinogens present in tobacco smoke, G to T transversions predominate at the *p53* locus in smoking-related tumours. UV light results in specific *p53* changes (transitions) in skin tumours (Nakazawa *et al.*, 1994). Other genes may exhibit these effects; for example, recently, methylation of the estrogen receptor in lung tumours has been observed to be frequent in plutonium-related tumours but low in tobacco-derived carcinogen-related tumours (Issa *et al.*, 1996).

Second, there is the relationship of somatic mutations to inherited defects. Knudson's (1983) findings in retinoblastoma are the paradigm for this type of relationship, with his observation that retinoblastoma was caused by two mutational events, one on each allele of the same gene. In the inherited form, one mutation is inherited via the germinal cells and a second occurs in a somatic cell due to an environmental insult or chance event. In the sporadic form, both mutations occur in somatic cells. The further observation that these events were accompanied by chromosome changes consistent with a loss of genetic material suggested a gene whose absence was required for cancer, a phenomenon termed 'anti-oncogene', but more commonly known today as tumour suppresser gene today. Many of the hereditary cancer syndromes listed in Table 1 have been shown to involve genes from this class, e.g. *p53* and *WT1*.

Recently, certain hereditary genes have been associated with somatic gene findings. Two recent examples include increased *p53* mutations in *GSTM1* null subjects with lung cancer (McGlynn *et al.*, 1995; Ryberg *et al.*, 1994), and, in a series of patients with colon cancer, rapid acetylators (*NAT2*) were more likely to exhibit K-*ras* gene mutations than intermediate or slow acetylators (Oda *et al.*, 1994).

Finally, somatic mutations are thought to indicate something about the degree of aggressiveness of the underlying cancer and may be related to stage, grade or clinical behaviour of the tumour. As observed by Boveri and many others, increased mutations are observed in advanced or poorly differentiated tumours. The presence or absence of chromosomal abnormalities predicts survival in chronic lymphocytic leukaemia (Criel *et al.*, 1997). Aneuploidy and other molecular or chromosome findings are related to bladder cancer aggressiveness (Walman *et al.*, 1991; Esrig *et al.*, 1994). An elaborate example of somatic mutations and tumour progression was provided by Vogelstein *et al.* (1988) with the demonstration of a series of molecular events characterizing progression from polyp to cancer to metastasis.

It should be appreciated that there are many types of somatic 'genetic' markers that involve the genetic material indirectly. A classic example is sister chromatid exchange, while telomere shortening (Sharma *et al.*, 1996) is a more recent

example. In each of these cases, study designs attempt to associate the finding with markers of exposure, susceptibility or effect (disease type, progression) using biomarkers. Study design issues and methodological consideration in this 'molecular epidemiology' approach have been described (Perera, 1982; McMichael, 1994).

Inborn errors of metabolism and the idea that genetically controlled influences on metabolism can determine disease

In 1908, Archibald Garrod described four rare recessively inherited conditions: albinism, alkaptonuria, cystinuria, and pentosuria. He postulated that these disorders were due to genetically based defects in normal biochemistry. He termed these 'inborn errors in metabolism' (Scriver *et al.*, 1995). This idea was a fertile concept for human disease etiology, initially for providing a genetic explanation for these conditions (perhaps the first description of an autosomal recessive disorder), but also for suggesting that biochemical diversity is the substrate upon which natural selection may act. Although the expression of genetic traits exhibits variability due to both environmental factors and the genotype, the inborn errors are highly penetrant in that virtually all those who inherit the trait exhibit features of the disease. This follows from the fact that the biochemical disorder involves endogenous metabolic processes, and therefore the consequence of the blocked pathway is unavoidable given the ubiquitous presence of the endogenous substrate (Scrivner *et al.*, 1995). A related class of phenomenon is distinguished from the inborn errors in that an inherited defect involves the aberrant metabolism of an extrinsic agent. The latter group are termed pharmacogenetic disorders (Kalow, 1962). This class of disorders is distinguished in that no disturbance is present in an individual with the trait until they are exposed to the particular medication dependent on the variant enzyme. Examples of this class of disorders are glucose-6-phosphate dehydrogenase (*G6PD*) deficiency, the porphyrias and malignant hyperthermia. Since carcinogens both require metabolic activation and are also subject to metabolic processes that facilitate elimination, the existence of variant versions of the enzymes that control these processes might be expected to alter susceptibility to the particular cancers.

Early studies of low penetrance genes and cancer in the general population

While the study of the highly penetrant 'single' genes would proceed through the approaches described earlier in cancer families, investigators looking for the more subtle effects of weaker 'susceptibility' genes would use different approaches, based on investigations set in the general population. There were four studies that established an interest in susceptibility to cancer and this class of genetic traits between 1973 and 1984. The first was the study of the phenotype of lymphocyte inducibility of (*CYP1A1*-dependent) aryl hydrocarbon hydroxylase activity (Kellermann *et al.* (1973). Three other studies in the 1980s were crucial. Ayesh (1984) studied debrisoquine metabolism (*CYP2D6*) in relation to lung cancer. Lower *et al.* (1979) examined the *N*-acetylation phenotype (*NAT2*) in aromatic amine-related bladder cancer using a sulfa probe drug, and Seidegard *et al.* (1986) examined glutathione-*S*-transferase mu activity (by measuring trans-stilbene oxidation in erythrocytes) (*GSTM1*) in relation to lung cancer. These studies had aspects that distinguished them from both prior and more modern studies. As described above, early cancer genetic studies were based on rare familial aggregations of cancer that were recognized clinically. The clinical phenotype was used to identify affecteds in order to initiate gene mapping studies, sometimes aided by a cytogenetic abnormality that would offer a clue to a gene location. In contrast, this new population-based approach selected a gene of interest (actually a phenotype, since there was no convenient method to identify the genotype at the time of these studies), generally based on a mechanistic hypothesis, i.e. bladder cancer should be more likely in individuals who are deficient in their genetically determined ability to inactivate aryl amines (i.e. bladder carcinogens). A population-based approach was required because the traits of interest derived from low penetrance genes. Thus, large numbers of study subjects were required to achieve the statistical power to detect the postulated differences in risk between the gene types. The phenotype was recognized, not by clinical findings (as in the inborn errors of metabolism), but rather through the use of a probe drug in two of the four studies (to determine a 'metabolic phenotype'), or in-vitro assay in the other two studies. In

Table 2. Contrasting the phenotype and genotype approaches to the characterization of pharmacogenetic variability in human population studies

Consideration	Phenotype	Genotype
Advantages	Historically validated approach Reflects physiological, in-vivo disposition of drug Reflects impact of inducers, inhibitors	Identifies hetero/homozygous subjects Technology evolving rapidly Simple, requires only germline DNA sample Invariant to illness, diet, medications, etc. Can be performed with micro-quantities Non-invasive samples (i.e. mouth wash, paraffin)
Disadvantages	Potentially distorted by numerous factors, i.e. drug–drug interaction Analysis typically more complex More patient cooperation required Phenotyping protocols poorly adapted to field study Invasive, time-consuming nature of test causes many exclusions	Functional status often requires study Some variants unknown Ethical questions arise since DNA may be used for other tests Allelic heterogeneity

contrast to the studies that characterized the inborn errors of metabolism (based on the recognition of a clinical phenotype), slow acetylators (e.g. of naphthylamine, bladder carcinogen) or poor metabolizers (of debrisoquine) exhibit no obvious clinical findings, and it is only upon challenge with the appropriate agent that differences in drug metabolism can be detected or clinical sequelae recognized. The probe drugs used to characterize the phenotype were of course innocuous non-carcinogens. While the phenotyping method of study is still used, modern studies increasingly depend upon direct identification of a genotype, and the advantages and disadvantages of each approach are summarized in Table 2.

In pharmacogenetic conditions, the phenotype does not always result in clinical sequelae, i.e. the condition is not fully penetrant. In pharmacogenetic conditions, acute sequelae result from specific exposures, typically to pharmaceutical agents (but also xenobiotics, carcinogens or endogenous compounds) that have some aspect of their metabolism dependent upon the enzyme (or receptor, immune factor or other element) that is subject to pharmacogenetic variability. The term 'ecogenetics' has been used to emphasize that agents that may be subject to this unusually high degree of metabolic variation are not limited to medications (Khoury *et al.*, 1988). Chronic conditions are thought to occur with altered frequency based on exposures over time to specific agents subject to this type of variability.

Over the last 10 years, many other genes have been studied in relation to various tumours. These have been reviewed recently (D'Erricco *et al.*, 1996). The genes of interest, mechanistic basis and phenotype/genotype approaches are listed in Table 3.

Broadening the proposed mechanism of low penetrance genes beyond metabolism

Early studies of cancer susceptibility focused on the idea that variation in enzymes made individuals activate or deactivate carcinogens differently, accounting for differences in susceptibility. For instance, cancer etiology, development, progression or prognosis may be influenced by hormones (breast and estrogen; prostate and androgen), infectious disease (*Helicobactor pylori* and gastric; hepatitis and liver cancer), and nutrient intake or immune factors. Examples of these mechanisms, and the genes that may influence them are listed in Table 4.

Methodological issues in studies of susceptibility genes in the population

Phenotype/genotype issues. With regard to low penetrance gene/cancer association studies, early phenotype-based work typically required the administration of a probe drug followed by the collection of a timed urine or blood sample. Given that it may be efficient to determine the status of a number of polymorphic traits at once, a 'cocktail' approach (i.e. administration of multiple probe drugs) has been used as data have shown that certain combinations of probes are safe and do not result in interference (Branch *et al.*, 1995). While phenotype approaches continue to be supplanted by genotype investigations, many new phenotypes with cancer associations (e.g. the bleomycin sensitivity assay and lung cancer, DNA repair assay and skin cancer) continue to be proposed (Cloos, 1996). It is therefore worthwhile to consider the advantages and disadvantages of each approach (Table 2) and the types of studies needed to validate these approaches.

Phenotyping studies: methodological issues. The difficulties involved in establishing a phenotyping approach as appropriate for population-based

Table 3. Polymorphic genes studied in relation to population cancers

Gene	Mechanism	Phenotype	Genotype	Hypothesized cancer
CYP1A1	Activates benzo[a]pyrene, role in estrogen metabolism (?)	AHH activity in lymphoblasts	MspI, exon 7 polymorphisms	Lung (Kellerman *et al.*, 1973), Breast (Taioli *et al.*, 1995)
CYP1A2	Activates heterocyclic amines, aryl amines	Caffeine breath test, caffeine urine, saliva metabolites	Under investigation	Bladder, colon (Lang *et al.*, 1994)
CYP2D6	Nicotine?	Debrisoquine, dextromethorphan metabolism	Inactivating and partially inactivating mutations exist	Lung (Ayesh *et al.*, 1984)
CYP2E1	Activates low molecular weight nitrosamines, alcohol-inducible		Polymorphisms studied but function poorly characterized	Nasopharyngeal carcinoma (Hildesheim *et al.*, 1995), lung (Uematsu *et al.*, 1991)
GSTM1	Detoxification of epoxides	Trans-stilbene oxide in RBCs	Absent gene	Lung (Seidegard *et al.*, 1986), bladder (Bell *et al.*, 1993)
GSTT1	Detoxification of ethylene oxide, butadiene	No	Absent gene	Myelodysplastic syndrome (Chen *et al.*, 1996)
NAT2	Detoxification of aromatic amines	Caffeine or sulfa metabolites	Mutations recognized	Bladder (occupational exposure-related) (Cartwright *et al.*, 1982), breast (Ambrosone *et al.* 1996)
NAT1	Detoxification of aromatic amines	?	Variants recognized	Gastric, bladder, colorectal (Bell *et al.*, 1995)

Table 4. Different categories of genes hypothesized to be involved in cancer susceptibility

Gene category	Gene example	Cancer site
Dominant oncogene	*ret, ras, myc*	Lung, others
Tumour suppresser genes	*p53, rb*	Lung, bladder, others (Wu *et al.*, 1995)
Hormones	Steroid metabolism	Prostate, breast, endometrium (Carey *et al.*, 1994)
Hormone receptor	Estrogen, progesterone and androgen receptor	Breast, endometrium, (Fuqua *et al.*, 1991), prostate (Irvine *et al.*, 1995)
Vitamins	Vitamin D, folic acid, B_{12}	Prostate (Taylor *et al.*, 1996), other medical illnesses (Jacques *et al.*, 1996)
Alcohol metabolism	*ADH, ALDH*	Oral cancer (also cirrhosis, alcoholism)
Nicotine metabolism	*CYP2D6, CYP2B6*	Smoking-related cancers (also COPD, others)
Addiction	Dopamine receptors	Smoking- and alcohol-related cancers (Comings *et al.*, 1994)
DNA repair	XP, X-linked lymphoproliferative syndrome, AT	Skin cancers, Burkitt lymphoma, others Hanawalt *et al.*, 1996
Immune function	*HLA*	Various rheumatological conditions

studies are substantial. As an example, we briefly review some work with the debrisoquine phenotype, a well-established example after two decades of work. It was discovered in the 1970s that after administration of debrisoquine, certain individuals (poor metabolizers) suffered prolonged hypotensive episodes. Population studies established that 10% of the general population was deficient in the enzyme required to hydroxylate the drug. A frequency histogram of phenotype values in a Western population exhibits a clear bimodal distribution. Family studies demonstrated that this deficiency was inherited (Evans *et al.*, 1980). The safety and feasibility of administering a tracer dose of debrisoquine had to be established in a clinical setting (Green-Gallo *et al.*, 1991). Circadian variation, the precise time necessary to establish a stable phenotype (number of hours of urine collection), recent diet, the influence of concurrent medications, and the comparability of day and night protocols to perform phenotyping were examined (Shaw *et al.*, 1990; Caporaso *et al.*, 1994). An issue that often clouds the interpretation of cancer association studies involving phenotypes is 'effect–cause' bias—i.e. could disease itself (or related treatment, nutritional status, etc.) alter the phenotype and bias study. Early studies showed that the debrisoquine metabolic ratio was unrelated to stage, grade or performance status in 92 patients with lung cancer, and a later, more definitive study showed that the metabolic ratio was invariant in individuals undergoing curative resection for lung cancer (Shaw *et al.*, 1994).

Assigning the phenotype. The result of debrisoquine phenotyping can be expressed as a ratio (the metabolic ratio, *mr*, is the molar ratio of parent drug to chief metabolite in the timed urine sample). This continuous variable must be assigned an ordinal category to specify a phenotype. One approach involves the use of a mixture model to analyse the overall distribution of *mrs* into phenotypic components. The resulting proportions may be compared with those expected under Hardy–Weinberg equilibrium. Parameter estimates for one-mix and two-mix models can be compared with a likeli-

hood ratio χ^2 test. This approach can be applied to assign cut-points for any pharmacogenetic study that involves the use of a probe drug (Caporaso *et al.*, 1989). As it has become more common for both phenotype and genotype information to be available, receiver operator curve approaches are an alternative to assign cut-points (DeLeo, 1993). Early studies of pharmacogenetic variation have relied exclusively on the phenotype, while advances in molecular biology have increasingly supplanted this approach with direct determinations of the genotype. Attractive features of genotyping include the avoidance of effect-cause bias and the attendant difficulties of phenotyping, and the possibility of identifying heterozygotes. While the relationship between the two study approaches is dynamic and evolving, some of the general strengths and weaknesses of each approach are indicated in Table 2.

Some general methodological issues involved in the selection of candidate low penetrance genes for population studies

A variety of factors can enter into the consideration of whether a proposed susceptibility factor is a worthy candidate for study. A list of questions should include the following:

- Is the trait stable in the individual? (Intra-individual variability should be small.)
- Does the trait vary in the general population? (Interindividual variation should be large.)
- Can diet, medications, circadian variation, disease or nutritional status distort phenotyping?
- Does the trait vary by ethnicity, age, gender, smoking, alcohol consumption, body weight, etc.? (Distorting factors should be understood so they can be controlled by design or adjusted for in the analysis.)
- Is the phenotype subject to induction or inhibition? If so, to what extent is the phenotype 'fixed' (i.e. subject to hereditary control), and to what extent do exogenous factors control it?
- Is the trait genetically controlled? (Have family studies demonstrated a mode of inheritance?)
- Does the gene have a plausible mechanism for involvement with cancer?
- Does the proposed substrate for the gene have any relation to cancer? (Is a procarcinogen activated, or carcinogen detoxified? Is there some other plausible mechanism, i.e. involving a nutrient, oncogene, receptor, hormone, etc.?)
- Are gene variants known, and do they have functional significance? Will all the important

Table 5. High and low penetrance gene studies

	High penetrance	Low penetrance
Role	Tend to be 'causal', i.e. necessary and sufficient to result in disease	Allele alters susceptibility, but is neither necessary nor sufficient for disease causation
Example	*Brca1* (breast/ovary) *APC* (polyposis coli) *RB* (retinoblastoma)	*CYP1A1* (lung) *CYP2D6* (lung) *GST-M1* (lung, bladder)
Gene type	Mutation	Polymorphism
Study setting	Family	General population or epidemiological studies
Strength of association	High (RR often > 200)	Low to moderate (RR, 2–10)
Relative or absolute risk	High	Low
Population attributable risk	Low	High
Gene–environment interaction	Secondary and variable	Primary and implicit
Role of environmental exposure	Secondary and variable	Critical

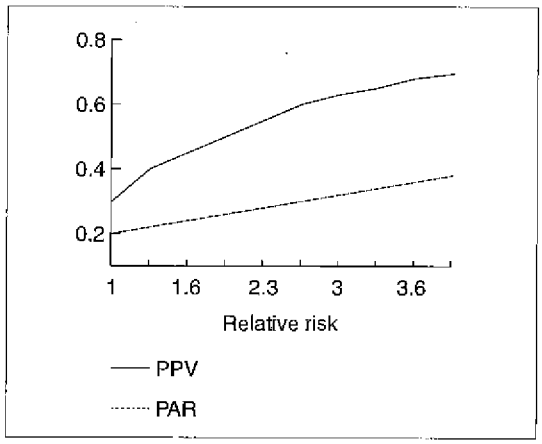

Figure 1. Positive predictive value (PPV) population attributable risk (PAR)for a 'common' susceptibility gene.

variants (polymorphisms) described in the population be studied?

- Is the gene known to act in a relevant organ? This may be the liver or the specific organ of the primary tumour.
- Have any studies in humans been performed? Have epidemiological studies in humans been carried out? If so, do the accumulated data support further study?

Summary of contrasts between high and low penetrance gene studies

There are certain contrasts between genes identified and studied in the context of families, i.e., the high penetrance 'single' genes listed in Table 1, and the low penetrance population-based studies suitable for the study of the susceptibility genes as described in Tables 2 and 5. These contrasts are summarized in Table 2. The 'rare' genes are associated with high relative and absolute risks but low attributable risk, while the common genes exhibit the opposite qualities, i.e. modest relative and absolute risks but large attributable risk.

There are contrasts between the implications of 'single' and 'susceptibility' genes for the individual and for public health. These are illustrated by a consideration of the positive predictive value (PPV, individual risk given a positive test) and the population attributable risk (or PAR, public health burden due to the disease, based on the combination of the at-risk genotype plus the relevant exposure) (Fig. 1). The calculations are based upon a method described by Khoury & James (1993) and are applied to a candidate susceptibility factor (*GSTM1*) studied in relation to lung cancer. PPV and PAR are calculated over a range of relative risks observed in published studies and assuming the following: the prevalence of relevant exposure is 35% (cigarette smoking). The risk of disease (lung cancer) is assumed to be non-zero in the absence of the gene. The relative risk of the exposure (tobacco smoking) is arbitrarily estimated at 10. A 'type 2 interaction', i.e. the gene does not result in increased risk in the absence of exposure (which is plausible, considering that lung cancer is rare in non-smokers), is assumed (Khoury & James, 1993).

The following conclusions are suggested from the graph. PAR is relatively high, while PPV is only modestly elevated, especially within the likely range of odds ratios, between 1 and 2. This suggests that this gene will have no role in screening or clinical testing and that clinical relevance to the individual is limited. The increase in the PAR, however, is significant, even with a modestly increased odds ratio. Because both the gene and the disease are common, even with the a 1.5 odds ratio, the burden of disease in the population, due to the gene (in the presence of exposure) is substantial.

References

Albertsen, H., Plaetke, R., Ballard, L., Fujimoto, E., Connolly, J., Lawrence, E., Rodriguez, P., Robertson, M., Bradley, P., Milner, B., Fuhrman, D., Marks, A., Sargent, R., Cartwright, P., Matsunami, N. & White, R. (1994a) Genetic mapping of the BRCA1 region on chromosome 17q21. *Am. J. Hum. Genet.*, 54, 516–525

Ambrosone, C.B., Freudenheim, J.L., Graham, S., Marshall, J.R., Vena, J.E., Brasure, J.R., Michalek, A.M., Laughlin, R., Nemoto, T., Gillenwater, K.A., Harrington, A.M., Shields, P.G. (1996) Cigarette smoking, *N*-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. *J. Am. Med. Assoc.*, 276, 1494–1501

Ayesh, R., Idle, J.R., Richie, J.C., Crothers, M.J. & Hetzel, M.R. (1984) Metabolic oxidation phenotypes as markers for susceptibility to lung cancer. *Nature*, 312, 169–170

Bell, D.A., Stephans, E.A., Castranio, T., Umbach, D.M., Watson, M., Deakin, M., Elder, J., Hendrickse, C., Duncan, H. & Strong, R.C. (1995) Polyadenylation polymorphism in the acetyltransferase 1 gene (NAT1) increases risk of colorectal cancer. *Cancer Res.*, 55, 3537–3542

Bell, D.A., Taylor, J.A., Paulson, D.F., Robertson, C.N., Mohler, J.L. & Lucier, G.W. (1993) Genetic risk and carcinogenesis exposure: common inherited defect of the carcinogen-metabolism gene glutathione S-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. *J. Natl Cancer Inst.*, 85, 1159–1164

Bodmer, W.F., Bailey, C.J., Bodmer, J. (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature*, 328, 614–616

Boveri, T.H., Zur Frage der Entstehung maligner Tumoren. *Jena*, 1914, pp. 25–41

Branch, R.A., Chern, H.D., Adedoyin, A., Romkes-Sparks, M., Lesnick, T.G., Persad, R., Wilkinson, G.R., Fleming, C.M., Dickinson, A.J., Sibley, G. (1995) The procarcinogen hypotheses for bladder cancer: activities of individual drug metabolizing enzymes as risk factors. *Pharmacogenetics*, 5, Spec. No. S97–102

Caporaso, N., Pickle, L.W., Bale, S., Ayesh, R., Hetzel, M. & Idle, J. (1989) The distribution of debrisoquine metabolic phenotypes and implications for the suggested association with lung cancer risk. *Genet. Epidemiol.*, 6, 517–524

Caporaso, N., Whitehouse, J., Monkman, S., Boustead, C., Issaq, H., Fox, S., Morse, M.A., Idle, J.R. & Chung, F.-L. (1994) In vitro but not in vivo inhibition of CYP2D6 by phenethyl isothiocyanate (PEITC), a constituent of watercress. *Pharmacogenetics*, 4, 275–280

Carey, A.H., Waterworth, D., Patel, K., White, D., Little, J., Novelli, P., Franks, S. & Williamson, R. (1994) Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum. Mol. Genet.*, 3(10), 1873–1876

Cartwright, R.A., Glashan, R.W., Rogers, H.J., Ahmad, R.A., Barham-Hall, D., Higgins, E. & Kahn, M.A. (1982) Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer. *Lancet*, 2, 842–845

Cavenee, W.K., Dryja, T.P., Philips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphee, A.L., Strong, L.C. & White, R.L. (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature*, 305, 779–784

Cavenee, W.K., Murphree, A.L., Shull, M.M., Benedict, W.F., Sparkes, R.S., Koch, E. & Nordenskjold, M. (1986) Prediction of familial predisposition to retinoblastoma. *New Engl. J. Med.*, 314, 1201–1207

Chen, H., Sandler, D.P., Taylor, J.A., Shore, D.L., Liu, E., Bloomfield, C.D. & Bell, D.A. (1996) Increased risk for myelodysplastic syndromes in individuals with glutathione theta 1 (*GSTT1*) gene defect. *Lancet*, 347, 295–297

Chissoe, S.L., Bodenteich, A., Wang, Y.F., Wang, Y.P., Bunan, D., Clifton, S.W., Crabtree J., Freeman, A., Iyer, K. & Jian, L. (1995) Sequence and analyses of the human ABL gene, the BCR gene, and region involved in the Philadelphia chromosome translocation. *Genomics*, 27, 67–82

Cloos, J. (1996) *Mutagen sensitivity: a phenotype reflecting susceptibility to head and neck squamous cell carcinoma.* Amsterdam

Comings, D.E., Muhleman, D., Ahn, C., Gysin, R. & Flanagan, S.D. (1994) The dopamine D2 receptor gene: a genetic risk factor for substance abuse. *Drug and alcohol dependence*, 34, 175–180

Criel, A., Verhoef, G., Vlietinck, R., Mecucci, C., Billet, J., Michaux, L., Meeus, P., Louwagie, A., Van Orshoven, A., Van Hoof, A., Boogiertes, M., van den Berghe, H., & DeWolf-Peeters, C. (1977) Further characterization of morphologically defined typical and atypical CLL; a clinical, immunophenoxy cytogenetic and prognostic study of 390 cases. *Br. J. Haematol.*, 97, 383–391

D'Erricco, A., Taioli, E., Xiang, C. & Vineis, P. (1996) Genetic metabolic polymorphisms and the risk of cancer: a review of the literature. *Biomarkers*, 1, 149–1731

Eckert, W.A., Jung, C. & Wolff, G. (1994) Presymptomatic diagnosis in families with adenomatosis polyposis using highly polymorphic dinucleotide CA repeat markers flanking the APC gene. *J. Med. Genet.*, 31, 442–447

Esrig, D.E., Elamjian, D., Groshen, S., Freeman, J.A., Stein, J.P., Chen, S.C., Nichols, P.W., Skinner, D.G., Jones, P.A. & Cote, R.J. (1994) Accumulation of nuclear p53 and tumor progression in bladder cancer. *N. Engl. J. Med.*, 331, 1259–1264

Evans, H.J. (1993) Molecular genetic aspects of human cancers: The 1993 Frank Rose Lecture. *Br. J. Cancer*, 68, 1051–1060

Evans, D.A.P., Mahgoub, A., Sloan, T.P., Idle, J.R. & Smith, R.L. (1980) A family and population study of the genetic polymorphism of debrisoquine oxidation in a white British population. *J. Med. Genet.*, 17, 102–105

Fearon, E.R., Cho, K.R., Nigro, J.M., Kern, S.E., Simons, J.W., Ruppert, J.M., Hamilton, S.R., Preisinger, A.C., Thomas, B., Kinkler K.W. *et al.* (1990) Identification of a chromosome 18q gene that is altered in colon cancer. *Science*, 247, 49

Fuqua, S.A.W., Hill, S.M., Chamness, G.C., Benedix, M.G., Greene, G.L., O'Malley, B.W. & McGuire, W.L. (1991) Progesterone receptor gene restriction fragment length polymorphism in human breast tumors. *J. Natl Cancer Inst.*, 83, 1157–1160

Gawin, B., Klamt, B., Konig, A., Thate, C., Le Paslier, D., Chumakov, I., Bhogal, R., Zehetner, G., Bruns, G. & Gessler, M. (1995) An integrated YAC clone for the

WAGR region on human chromosome 11p13-p14.1. *Genomics*, 30, 37–45

Green-Gallo, L.A., Buivys, D.M., Fisher, K.L., Caporaso, N., Slawson, R.G., Elias, G., Didolakar, M.S., Ivusich, W.I. & Resau, J.H. (1991) A protocol for the safe administration of debrisoquine in biochemical epidemiologic research protocols for hospitalized patients. *Cancer*, 68, 206–210

Haas, O.A., Argyriou-Tirita, A. & Lion, T. (1992) Parental origin of chromosomes involved in translocation t(9,22). *Nature*, 359, 414–416

Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. & King, M.C. (1990) Linkage of early-onset beast cancer families to chromosome 17q21. *Science*, 250, 1684–1689

Hanawalt, P.C. (1996) Role of transcription-coupled DNA repair in susceptibility to environmental carcinogenesis. *Environ. Health Perspect.*, 104, Suppl. 3, 547–551

Hilldesheim, A., Chen, C.J., Caporaso, N.E., Cheng, Y.J., Hoover, R.N., Hsu, M.M., Levine, P.H., Chen, I.H., Chen, J.Y., Yang, C.S., Daly, A.K. & Idle, J.R. (1995). Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: results from a case-control study conducted in Taiwan. *Cancer Epidemiol. Biomarkers Prev.*, 4, 607–610

Hussussion, C.J., Struewing, J.P., Goldstein, A.M., Higgins, P.A.T., Ally., D.S., Stratan, M.D., Clark, W.H., Jr, Tucker, M.A. & Dracopoli, N.C. (1994) Germline p16 mutations in familial melanoma. *Nature Genet.* 8, 15–21

Irvine, R.A., Yu, M.C., Ross, R.K. & Coetzee, G.A. (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res.*, 55, 1937–1940

Issa, J.-P.I., Baylin, S.B. & Belinsky, S.A. (1996) Methylation of the estrogen receptor CpG island in lung tumors is related to the specific type of carcinogen exposure. *Cancer Res.*, 56, 3655–3658

Jacques, P.F., Bostom, A.G., Willaims, R.R., Ellison, R.C., Eckfeldt, J.H. *et al.* (1996) Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations (1996). *Circulation*, 93(1), 7–9

Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bose, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E.H. & Scott, M.P. (1996) Human homolog of patched a candidate gene for the basal cell nevus syndrome. *Science*, 272, 1668–1671

Kalow, W. (1962) Pharmacogenetics: heredity and response to drugs. Philadelphia, W.B. Saunders

Kamb, A., Shattuck-Eidens, D., Eeles, R., Liu, Q., Gruis, N.A., Ding, W., Hussey, C.,Tran, T. Miki, Y. & Weaver-Feld, . (1994) Analysis of p16 gene (CDKN2) as a candidate locus for the 9p melanoma susceptibility locus. *Nature Genet.*, 8, 23–26

Kellerman, G., Shaw, C.R. & Luyten-Kellerman, M. (1973) Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. *New Engl. J. Med.*, 289, 934–937

Khoury, M.J., Adams, Jr, M.J. & Flanders, W.D. (1988) An epidemiologic approach to ecogenetics. *Am. J. Hum. Genet.*, 42, 89–95

Khoury, M.J. & James, L.M. (1993) Population and familial relative risks of disease associated with environmental factors in the presence of gene-environmental interaction. *Am. J. Epidemiol.*, 137(11), 1241–1250

Kinzler, K.W., Nilbert, M.C., Vogelstein, B., Bryan, T.M., Levy, D.S., Smith, K.J., Preisinger, A.C., Hamilton, S.R., Hedge, P., Markham, A. *et al.* (1991) Identification of a gene at chromosome 5q21 that is mutated in colorectal cancer. *Science*, 251, 1366–1370

Knudson A.G. (1993) All in the (cancer) family. *Nat. Genet.*, 5, 103–104

Knudson, A.G., Hethcote, H.W. & Brown, B.W. (1975) Mutation and childhood cancer: a probabilistic model for the incidence of retinoblastoma. *Proc. Soc. Natl Acad. Sci. USA*, 72, 5116

Knudson, A.G. (1983) Model hereditary cancers of man. *Prog. Nucl. Acid Res. Mol. Biol.*, 29, 17–25.

Lang, N.P., Butler, M.A., Massengill, J., Lawson, M., Stotta, R.C., *et al.* (1994) Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclc amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiol. Biomarkers Prev.*, 3, 675–682

Lee, W.H., *et al.* (1979) Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science*, 235, 1394

Lohmann, D., Horsthemke, B., Gillessen-Kaesbach, G., Stefani, F.H. & Hofler, H. (1992) Detection of small RB1 gene deletions in retinoblastoma by multiplex PCR and high resolution gel electrophoresis. *Hum . Genet.*, 89(1), 49–53

Lower, G.M., Nilsson, T., Nelson, C.E., Wolf, H., Gamsky, T.E. & Bryan, G.T. (1979) N-Acetyltransferase phenotype and risk in urinary bladder cancer: Approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. *Environ. Health Perspec.*, 29, 71–79

Malkin, D., Li, F.P., Stong, L.C., Fraumeni, J.F. Jr, Nelson, C.E., Kim, D.H., Kassel, HJ., Gryka, M.A., Bischoff, F.Z., & Tainsky, M.A. (1990) Germline p53 mutations in a famil-

ial syndrome of breast cancer, sarcoma, and other neoplasms. *Science*, 250, 1233–1238

McGlynn, K.A., Rosvold, E.A., Lustbader, E.D., Hu, Y. *et al.* (1995) Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. *Proc. Natl Asoc. Sci., USA*, 92, 2384–2387

McMichael, A.J. (1994) "Molecular epidemiology" – new pathway or new traveling companion. *Am. J. Epidemiol.*, 140, 1–11

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M. & Ding, W. et al.(1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science*, 266, 66–71

Mulligan, L.M., Kwok, J.B., Healey, C.S., Elsdon, M.J., Eng., C., Gardner, E., Love, D.R., Mole, S.E., Moore, J.K, Papi, L. (1993) Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature*, 363, 458–460

Nagase, H., Miyoshi, Y, Horii, A., Ogawa, M., Utsunomiyi, J., Baba, S., Sasazuki, T. & Nakamura, Y. (1992) Correlation between the location of germline mutations in the APC gene and the number of colorectal polyps in familial adenomatosis polyposis patients. *Cancer Res.*, 52, 4055–4057

Nakazawa, H., English, D., Randell, P.L., Nakaema, K., Matel, N., Armstrong, B.K., Yamasaki, H. (1994) UV and skin cancer: specific p53 mutations in normal skin as a biologically relevant exposure measurement. *Proc. Natl Acad. Sci., USA*, 91, 360–364

Nicolaides, N.C., Papadopoulos, N., Wei, Y.F. (1994) Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature*, 371, 649

Nowell P.C. & Hungerford, D.A. (1960). A minute chromosome in human chronic granulocytic leukemia. *Science*, 132, 1497

Nystrom-Lahti, M., Parsons, R., Sistonen, P., Pylkkanen, L., Aaltonen, L.A., Leach, F.S., Hamilton, S.R., Watton, P., Bronson, E., Fusaro, R., *et al.* (1994) Mismatch repair genes on chromosome 2p and 3p account for a major share of hereditary nonpolyposis colorectal cancer families evaluable by linkage. *Am. J. Hum. Genet.*, 55, 659–665

Oda, Y., Tanaka, M. & Nakanishi, I. (1994) Relation between the occurrence of K-ras gene point mutations and genotypes of polymorphic N-acetyltransferase in human colorectal cancer carcinmomas. *Carcinogenesis*, 15 (7) 1365–1369

Ohyashiki, J.H., Ohyashiki Km Fujimura, T., Kawakubo, K., Shimamoto, T., Iwabuchi, A. & Toyama, K.(1994) Telomere shortening associated with disease patterns in myelodysplastic syndromes. *Cancer Res.*, 54, 3557–3560

Ott, J. (1991) *Analysis of human genetic linkage.* Baltimore, MD, The Johns Hopkins University Press

Papadopoulos, N., Nicolaides, N.C., Wei, Y.F., Ruben, S.M., Carter, K.C., Rosen, C.A. Haseltine, \W.A., Fleishmann, R.D., Fraser, C.M. & Adams, M.D. (1994) Mutation of a mutL homolog on hereditary colon cancer. *Science*, 263, 1625–1629

Perera F.P. (1982) Molecular cancer epidemiology: a new tool in cancer prevention. *J. Natl Cancer Inst.*, 78, 887–898

Rowley, J. Chromosome abnormalities in cancer. (1980) *Cancer Genet. Cyto.*, 2, 175

Ryberg, D., Kure, E., Lystad, S., Skaug, V., Strageland, Mercy, I., Borresen, A.L. & Haugen, A. (1994) p53 mutations in lung tumors: relationship to putative susceptibility markers for cancer. *Cancer Res.*, 54, 1551–1555

Savitsky, K., Bar-shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S. Uziel, T., Sfez, S. et al (1995) A single ataxia telangiectasia gene with a product similar to PI-3 Kinase. *Science*, 268, 1749–1753

Scrivner, C.R., Beaudet, A., Valle, D. & Sly, W.S. (eds) (1995)The metabolic and molecular basis of inherited disease, Vols I–III, 7th edn. McGraw-Hill, NY.

Seidegard, J., Pero, R.W. Miller, D. & Beattie, E.S. (1986) A glutathione transferase in human leukocytes as a marker for susceptibility to lung cancer. *Carcinogenesis*, 7, 751–753

Seizinger, B.R., Rouleau, G., Ozelius, L.J., Lane, A.H., St George-Hyslop, P., Huson, S., Gusella, J.F., Martuza, R.L. (1987) Common pathologic mechanism for three tumor types in bilateral acoustic neurofibromatosis. *Science*, 236, 317–319

Sharma, H.W., Maltese J.-Y., Zhu, X., Kaiser, H.E. & Narayanan, R. (1996) Telomeres, telomerase and cancer: is there a magic bullet? *Anticancer Res.*, 16, 511–516

Shaw, G.L., Falk, R.T., Caporaso, N.E., Issaq, H.J., Kase, R.G., Fox, S.D. & Tucker M.A. (1990) Effect of diurnal variation on debrisoquine metabolic phenotyping. *J. Natl Cancer Inst.*, 82, 1573–1576

Shaw, G.L., Falk, R.T., Deslavriers, J., Nesbitt, J.C., Frame, J.N. & Hoover, R.N. (1994) Lung tumor resection does not affect debrisoquine metabolism. *Cancer Epidemiol. Biomarkers Prev.*, 3, 141–144

Soini, Y., Chia, S.C., Bennett, W.P., Groopman, J.D. , Wang, J.S., DaBenedetti, V.M., Cawley, H., Welsh, J.A., Hasen, C., Bergasa, N.V., Jones, E.A., DisBisceglie, A.M.,

Trivers, G.L., Sandoval, C.A., Calderon, I.E., Munoz Espinoza, L.E. & Harris, C.C. (1996) An aflatoxin-associated mutational hotspot at codon 249 in the p53 tumorspressor gene in HCC. *Cancer*, 17(5), 1007–1020

Sozzi, G., Veronese, M., Negrini, M., Baffa, R., Cotticelli, M.G., Inoue, H., Tornelli, S., Pilotti, S., De Gregorio, L., Pastorino, U., Pierotti, M.D., Ohta, M., Huebner, K. & Croce, C.M. (1996) The *FHIT* gene at 3p14.2 is abnormal in lung cancer. *Cell*, 85, 17–26

Sparkes, R.S., Murphree, A.L. & Lingua, R.W. (1983) Gene for hereditary retinoblastoma assigned to human chromosome 13 linkage to esterose-D. *Science*, 219, 971–973

Taioli, E., Trachman, Chen, X., Toniolo, P. & Garte, S. (1995) A CYP1A1 restriction fragment length polymorphism is associated with breast cancer in African-American women. *Cancer Res.*, 55, 3757–3758

Tanaka, K., Maoyuki, M., Satokata, I., *et al.* (1990) Analysis of a human DNA excision repair gene involved in group A XP containing a zinc-finger domain. *Nature*, 348, 73–76

Taylor, J.A., Hirvonen, A., Watson, M., Pittman, G., Mohler, J.L. & Bell, D.A. (1996) Association of prostate cancer with vitamin D receptor gene polymorphism. *Cancer Res.*, 56, 4108–4110

Thakker, R.V., Wooding, C. & Pang, J.T. (1993) Linkage analysis of 7 polymorphic markers at chromosome IIpII.2-11ql3 in 27 multiple endocrine neoplasms type I families. *Ann. Hum. Genet.*, 57, 17–25

Uematsu, F., Uckuchi, H., Motomiya, M., Abe, T., Sagami, I., Ohmachi, T. *et al.* (1991) Association between restriction fragment length polymorphism of the human cytochrome P450IIE1 gene and susceptibility to lung cancer. *Jpn J. Cancer Res.*, 82, 254–256

Upadhya, M., Maynard, J., Osborn, M., Huson, S.M., Ponder, M., Ponder B.A. & Harper, P.S. (1995) Characterisation of germline mutations in the neurofibromatosis type 1 (NF1) gene. *J. Med. Genet.* 32, 706–710

Vogelstein B., Fearon, E.R. & Hamilton, S.R. (1988) Genetic alterations during colorectal tumor development. *N. Engl. J. Med.*, 319, 535–532

Walman, F.M., Carroll, P.R. & Kerschmann, R. (1991) Centromeric copy number of chromosome 7 is strongly correlated with tumor grade and labeling index in human bladder cancer. *Cancer Res.*, 51, 3807–3813

Wertalecki, W., Rouleau, G.A., Superneav, D.W., Forehand, L.W., Williams, J.P., Haines, J.L., Gusella, J.F. (1988) Neurofibromatosis 2: clinical and DNA linkage studies of a large kindred. *N. Engl. J. Med.*, 319 (S), 278–283

Whang-Peng, J.W.P., Kao-Shan, C.S., Lee, E.C., & Bunn, P.A., Carney, D.N., Gazdar, A.F. & Minna, J.D. (1982) Specific chromosome defect associated with small-cell lung cancer: deletion of 3p(14–23). *Science*, 215, 181–182

Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D. *et al.* (1994) Localization of the breast cancer susceptibility gene, BRCA2 to chromosome 13q12–13. *Science*, 266, 2088–2090

Wu, W.-J,, Kakehi, Y., Habuchi, T., Kinoshita, H., Ogawa, O., Terachi, T., Huang, C.-H., Chiang, C.P. & Yoshida, O. (1995) Allellic frequency of p53 gene codon 72 polymorphism in urologic cancers. *Jpn J. Cancer Res.*, 86, 730–736

Yokota, J., Wada, M., Shimosato, Y. *et al.* (1987) Loss of heterozygosity on chromosomes, 3,13,17 in small cell lung carcinoma and on chromosome 3 in adenocarcinoma of the lung. (1987) *Proc. Natl Acad. Sci. USA*, 84, 9252–9256

Zuo, L., Weger, J., Yong, Q., Goldstein, A.M., Tucker, M.A., Walker, G.J., Hayward, N. & Dracopoli, N.C. (1996) Germline mutations in the $p16^{INXYa}$ binding domain of CDK4 in familial melanoma. *Nature Genet.* 12, 97–99

Corresponding author

N. Caporaso
Genetic Epidemiology Branch
National Cancer Institute, EPN 439
6130 Executive Blvd., Rockville, MD 20892, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Gene – environment interactions in the application of biomarkers of cancer susceptibility in epidemiology

S. Garte, C. Zocchetti and E. Taioli

Metabolic susceptibility genes are important determinants of individual susceptibility to the effects of environmental carcinogens. These genes follow the form of 'type 2' gene–environment interaction, whereby the polymorphic genetic risk factor functions only in the presence of an environmental exposure. Two different effects of carcinogen dose have been observed for these genes. Sometimes, increasing dose leads to a decreasing interaction, so that cases with the genetic risk factor have lower exposures than those cases without it. Other examples of a direct dose effect, whereby increasing exposure leads to increased interaction, have also been described. We propose a model based on multiple logistic regression to assess the nature of the dose effect in this type of gene–environment interaction. This model allows for distinction between these two dose effects, and other effects such as protective or non-interactive effects of environmental and genetic risk factors.

It has been known for many decades that human diseases are caused by some combination of environmental and genetic factors. The relative influence of the two is obviously quite variable and forms a spectrum with certain highly infectious diseases at one end and severe genetic diseases of metabolism at the other. For the great majority of human diseases (excluding extreme cases such as Ebola fever or Down syndrome), purely environmental or purely genetic etiologies are insufficient to explain individual variability in occurrence, prognosis or outcome. This is especially the case with the most important chronic diseases of modern industrialized society, including heart disease and cancer (Hirayama, 1989; Hegele, 1992; Lane *et al.*, 1992; Hunt *et al.*, 1993; Tiret *et al.*, 1993; Brennan & Silman, 1994; Hayden *et al.*, 1994; Hwang *et al.*, 1995). For these categories of disease, a great number of environmental and genetic risk factors have been identified, and it is probably safe to say that for all cases of cancer both types of factors must play some role in disease causation. In general, when one assumes multiple causes, it is useful to ask whether interactions between these independent causes exist. The issue of interaction between the environment and genetic factors is an old one in biology, and in a classic paper entitled 'The interaction of nature and nurture' published in 1946; the great geneticist Haldane (1946) discussed gene–environment interactions in disease causation. Since then, a substantial literature has been produced on the interaction between genetic and environmental agents, and even a review of the review literature would not be feasible here. This paper will instead be limited to a discussion of the models of gene–environment interactions that are of primary importance in the area of individual human susceptibility to the carcinogenic effects of environmental carcinogens. With the introduction of new biomarkers of cancer susceptibility, new paradigms for the classical terms of interaction and confounding are needed (Hulka, 1991). The focus will be on cancer as the disease end-point, and in particular on the most common (usually termed sporadic) category of cancer in the human population, as opposed to cancers such as retinoblastoma, Wilms' tumour and other examples of purely or mostly inherited cancers (which also include the familial cancers of the breast, colon and brain).

There is a general consensus that the majority of cases of human cancer (sporadic tumours of the lung, bladder, breast, etc.) are of largely environmental origin, where environment is defined broadly as including lifestyle and exposures to agents contained in cigarette smoke and diet, as well as occupational or environmental exposure to

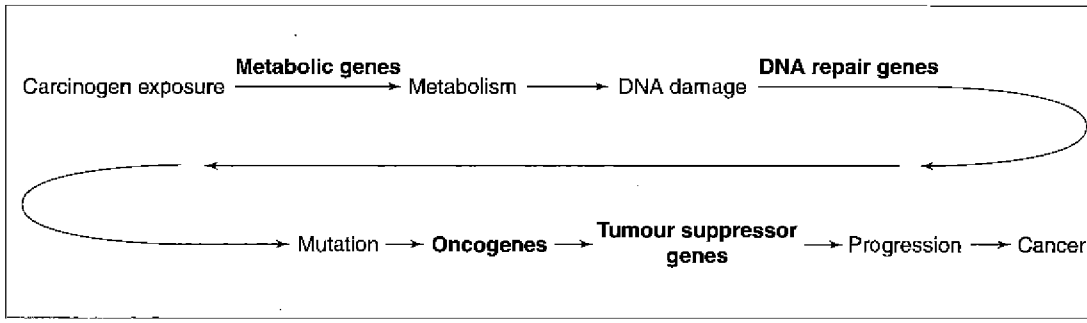

Figure 1. Mechanistic pathway to cancer.

carcinogens. Evidence supporting a dominant environmental etiology of human cancer comes from decades of epidemiological research on migrant populations and other studies (Higginson, 1980; Doll & Peto, 1981; Garte, 1992). Despite the importance of environmental exposure in human cancer, the evidence for some form of genetic influence on almost all cancer etiology is quite compelling. Unlike most strains of laboratory animals used in carcinogenesis research, human beings are highly outbred and genetically diverse. They also show an extremely broad range of phenotypic responses to environmental stimuli. For example, there is a wide variation in human response to specific carcinogenic exposures, cigarette smoking being one of several examples (Sellers *et al.*, 1992). This interindividual variability in response is seen not only in disease end-points but also when one uses biomarkers of exposure or biological effect such as DNA adducts or others (Marquis & Siek, 1988; Harris, 1989; Perera *et al.*, 1991). While such variability presents formidable obstacles to the development of certain biomarkers, it also presents an exciting opportunity for research into the mechanisms behind the variability and the development of markers of susceptibility in individual humans.

Genes and human cancer

It is vitally important at this stage to distinguish between different types of genes that are involved in human carcinogenesis. One broad category of cancer genes includes both the dominant acting oncogenes, such as *ras* and *myc*, and the recessive tumour suppressor genes such as *p53* and *Rb*. These genes exert their effects as part of the biological pathway leading to tumorigenesis. There is ample evidence that environmental carcinogens, from radiation to cigarette smoke, interact either directly or indirectly with the structure and/or function of oncogenes and tumour suppressor genes. Somatic mutations in these genes caused by exposure to chemicals and radiation are of critical importance in the etiology of sporadic human cancer. Such interactions may represent some of the most important mechanistic pathways in human carcinogenesis. The results of these interactions, such as mutations or gene deletions, have been proposed as biomarkers of biological effect for certain carcinogens, such as aflatoxin (Hollstein *et al.*, 1993) and ionizing radiation (Vahakagas *et al.*, 1992). However, even in the absence of such interactions, these genes may still play an important role in cancer causation if activating or inactivating mutations occur in the germline, such as in the case of retinoblastoma or Li-Fraumeni syndrome (Friend *et al.*, 1988; Frebourg & Friend, 1992), or in the case of rare H-*ras* alleles (Conway *et al.*, 1995).

Another category of genes that are important in human carcinogenesis does not play any direct role in the mechanistic pathway leading from carcinogen exposure to cancer; instead these genes have an influence on events that occur on the pathway. Genes involved in DNA repair (Lehman *et al.*, 1992; Samson, 1992; Smith *et al.*, 1994a) are an example of this category, as are genes that mediate the conversion of carcinogenic chemicals to their ultimate active forms. The distinction between these two types of 'cancer genes' with respect to their position on the mechanistic pathway to cancer is illustrated in Fig. 1. Genes responsible for the repair of DNA damage by carcinogens have been intensively studied because of their importance in carcinogenic mechanisms. They are certainly participants in one type of gene–environment interaction.

However, they have not been widely used to date as biomarkers of either biological effect or susceptibility. This situation is likely to change with the recent cloning of the gene for ataxia telangiectasia (AT) (Savitsky *et al.*, 1995). Carriers of the homozygous variant of this gene have greatly increased sensitivity to the carcinogenic effects of ionizing radiation. While homozygotes (who show the extreme symptoms of the disease) are very rare, it has been estimated that the frequency of heterozygotes who may also have some degree of increased susceptibility to radiation-induced carcinogenesis may be as high as 10% of the population. Further study of this and other repair-related genes as biomarkers of cancer susceptibility, and in relation to interaction with environmental factors, should provide important and useful information in the future.

The focus of this report will be on a number of genes that have been studied for many decades, after the initial discovery by E.C. and J.A. Miller (Conney *et al.*, 1956) that most carcinogens undergo complex metabolic reactions and that the actual carcinogenic compounds responsible for DNA binding or damage are usually electrophilic metabolites of the parent compound to which the organism is exposed. As shown in Fig. 1, these genes act as effect modifiers on the carcinogenic pathway from exposure to cancer.

Metabolic genes

A number of genes involved in the metabolism of carcinogens have been shown to play a role in the risk of certain human cancers. In most cases, the putative biochemical mechanism by which such genetic factors exert their effects is fairly straightforward and is related to the actual dose of the active carcinogenic metabolite that reaches the genome in the target cell. Increased metabolism of a carcinogenic precursor to the ultimate carcinogen and loss of function of a conjugation mechanism for elimination of the active metabolite are examples of how alterations in the activity of certain gene products can affect the biologically relevant dose in different individuals, even when exposures are equivalent. Reviews of metabolic susceptibility genes have been published (Idle *et al.*, 1992; Daly *et al.*, 1994; Hirvonen, 1995; Raunio *et al.*, 1995; Rothman, 1995; Vineis, 1995), and there is a growing literature on the use of these genes (first by phenotype and later by genotype analysis) as biomarkers of human cancer susceptibility. By definition, these genes function only in the context of interaction with the environment, since the substrates of their gene products are xenobiotic chemicals or their metabolites.

The metabolism of carcinogens, like that of most toxic agents, generally proceeds through two phases. (Garte & Kneip, 1988). In the first, unreactive, non-polar compounds are converted, usually by oxidation reactions, to electrophilic highly reactive intermediates. These are then able to form complexes with conjugating molecules such as glucose or glutathione in phase 2 conjugation reactions. The complex is usually harmless and easily excreted. However, the electrophilic metabolite may be able to react with other cellular nucleophiles such as DNA before conjugation can occur. This is often the first step in the initiation of a carcinogenic process.

Table 1 presents a summary of the most commonly studied metabolic susceptibility genes. The cytochrome P450 enzymes, which represent a large multigene family with differing substrate specificities, are important in phase 1 reactions. The *CYP1A1* gene product, aromatic hydrocarbon hydroxylase (AHH), for example, catalyses the first oxidative step in the metabolism of polycyclic aromatic hydrocarbons, such as those found in tobacco smoke, to carcinogens. An *Msp*I restriction fragment length polymorphism (RFLP) in the 3′ non-coding region of *CYP1A1* was found to be associated with AHH enzymatic activity in a family study (Petersen *et al.*, 1991). A second polymorphism was found in the catalytic region (exon 7) of the *CYP1A1* gene, closely linked to the *Msp*I RFLP (Hayashi *et al.*, 1991; Hirvonen *et al.*, 1992; Cosma *et al.*, 1993b). Both polymorphisms have been associated with lung cancer in Japanese populations but not in Caucasian (Kawajiri *et al.*, 1990; Tefre *et al.*, 1991; Hirvonen *et al.*, 1992; Shields *et al.*, 1993) or African-American (Shields *et al.*, 1993; Sugimura *et al.*, 1994) populations. This discrepancy may reflect the different frequencies of the two polymorphisms between ethnic groups, as we previously reported (Cosma *et al.*, 1993b; Taioli *et al.*, 1995b). We have recently found a very strong association between breast cancer risk and the homozygous *Msp*I RFLP in African-American women (Taioli *et al.*, 1995c). We have also described a third

Table 1. Examples of metabolic susceptibility genes

Gene	Metabolic pathway	Cancer sites
GSTM1	Conjugation of organic epoxides with reduced glutathione	Lung, bladder, colon, stomach, breast, liver
CYP2D6	Hydroxylation of lipophilic xenobiotics, possibly NNK	Lung, bladder, breast
NAT2	*N*-Acetylation of arylamines and *N*-hydroxylated heterocyclic arylamines	Bladder, lung, colorectal, breast
CYP1A1	Metabolism of polycyclic aromatic hydrocarbons, TCDD and estrogens	Lung, stomach, colon, breast
CYP2E1	Oxidation of *N*-nitrosamines, alcohol	Lung, bladder, colon

polymorphism in the human *CYP1A1* gene, an African-American-specific *Msp*I RFLP in intron 7 (Crofts *et al.*, 1993). This RFLP, which has not been detected in over 300 Caucasians nor in Asians, occurs in 15% of African-Americans and is associated with an increased risk of adenocarcinoma of the lung (Taioli *et al.*, 1995a).

Other phase 1 genes that have been identified as susceptibility factors are *CYP1A2* (Sinha *et al.*, 1994; Catteau *et al.*, 1995), a gene that is induced by heterocyclic aromatic hydrocarbons; *CYP2D6*, which is responsible for the metabolism of the drug debrisoquine; and *CYP2E1* (Stephens *et al.*, 1994; Kato *et al.*, 1995; Watanabe *et al.*, 1995). Phenotype analysis suggested a role for the debrisoquine polymorphism as a susceptibility factor in human lung cancer, but this has become less clear as genotype data has been gathered. The specific carcinogen associated with the *CYP2D6* activity and the biochemical effect of the polymorphism are not known. The *CYP2E1* gene is inducible by ethanol, and is involved in the metabolism of such carcinogens as butadiene, benzene and carbon tetrachloride.

Two phase 2 genes have received wide attention as metabolic susceptibility markers. The gene coding for one form of glutathione *S*-transferase, *GSTM1*, is missing in about one-half of Caucasians (Hirvonen *et al.*, 1993; Alexandrie *et al.*, 1994; Ichiba *et al.*, 1994; Kihara *et al.*, 1994; Nakajima *et al.*, 1995). This null allele results in a lower level of glutathione conjugation of PAH metabolites and possibly other carcinogenic substances, and has been associated with increased risk of lung cancer in several studies of Caucasians and Asians. Another gene first identified as a metabolic susceptibility marker through early phenotype analysis is *N*-acetyl transferase or *NAT2* (Meyer 1994; Vineis *et al.*, 1994; Yu *et al.*, 1994). A homozygous polymorphism in this gene renders individuals slow acetylators of certain substrate drugs such as isozionid. The *NAT2* gene participates in a complex metabolic web, and individuals with the polymorphism (slow acetylators) may be at higher risk for bladder cancer from carcinogenic exposure to arylamines, but at lower risk for colon cancer. This complication is not surprising given the tissue specific complexity of metabolic pathways and competing reactions catalysed by a number of genes.

As discussed by Caporaso (1996), there are major differences between metabolic gene polymorphisms and inherited mutations in cancer genes of the first category (such as *BRCA1*, for example). These differences have profound implications for screening, prevention, public health and risk assessment. For example, polymorphisms in metabolic genes tend to be much more common in the population (from 5 to 50%) than mutations in cancer genes. At the same time, the increased cancer risks associated with metabolic gene polymorphisms are usually on the order of twofold compared to very high odds ratios for inherited mutations in tumour suppressor genes or oncogenes. Perhaps

the most important difference relates to the clinical significance of these inherited genetic risk factors. It is not clear how a physician can counsel a person who is found to contain a genetic predisposition for cancer (such as is the case in Li-Fraumeni, etc.) independently of any exposure. On the other hand, since metabolic gene polymorphisms specifically confer increased sensitivity to the effects of environmental carcinogens, increased surveillance and avoidance of such agents may be an effective strategy for cancer prevention for individuals carrying such polymorphisms.

Types of gene–environment interactions

A number of authors have discussed various biological forms of gene–environment interaction as applied to epidemiological studies. M.J. Khoury and his co-workers, stressing the population and family genetics aspect of such interactions and their implications for public health, have described six types of gene–environment interaction (GEI) (Khoury *et al.*, 1988, 1993, 1995; Khoury & James, 1993; Khoury & Wagener, 1995). In the first type, neither the environmental exposure nor the genetic risk factor (GRF) have any effect by themselves in the absence of the other, but when both are present, interaction between them causes disease. The example given is that of phenylalanine exposure (in the diet) and phenylketonuria genotype. This type of GEI is rare and not important in human carcinogenesis. A type 2 GEI is defined as one in which the GRF has no effect on disease in the absence of the relevant exposure, but which can function to exacerbate the effects of the exposure. In this type, the exposure by itself increases the risk of disease, even in the absence of the GRF. This is the most important type of GEI for human carcinogenesis related to metabolic susceptibility genes and we will return to it shortly. The third type is the converse of the second, in that the GRF can produce disease in the absence of exposure; exposure mediates the effect of the GRF but, without the GRF, has no role in disease etiology. While this type of interaction may be important in certain cases of human carcinogenesis, not enough is known as yet regarding the detailed mechanisms by which specific carcinogens act to be able to say that any environmental carcinogen has effects only on those people with a particular genetic make-up. It is certainly possible, however, that future research will render this type of GEI highly significant for human cancer. The fourth type described by Khoury is common and important in cancer, and occurs when both the exposure and the GRF carry some risk for disease, but the combination is interactive and/or synergistic. The cancers associated with the DNA repair gene deficiencies such as AT (see above) or xeroderma pigmentosum (a repair gene deficiency disease associated with exposure to UV radiation) are examples. Most of the first category of cancer genes (*c-myc*, *p53*, etc.) belong to this type of GEI, since the gene mutations themselves carry an increased risk of cancer which is exacerbated by exposure to environmental carcinogens (which are still carcinogenic in the absence of such gene mutations). The latter two types of GEI described by Khoury refer to cases in which the GRF is protective.

Ottman (1990, 1994, 1995) has also described five similar types of gene–environment interaction. In the first, the disease may be caused by either the genetic or the environmental agent, but the genotype increases the expression of the agent. The second and third are the same as those described by Khoury. In the fourth type, both environmental and genetic risk factors must be present to cause the disease, equivalent to Khoury's type 1. In the final model, both factors influence risk by themselves, but with an interaction between them.

The critical point made by both groups is that the term interaction covers a variety of biological phenomena involving gene products and xenobiotics. The specific form of the gene–environment interaction is clearly as important as the fact that such interaction exists. Type 2 interaction (as used by both R. Ottman and M.J. Khoury) is the most relevant to GEI related to metabolic susceptibility genes and human carcinogenesis. In this type, the cancer is caused by exposure to an environmental agent. If there is no exposure, then the presence or absence of the genetic risk factor is irrelevant for disease causation. Only when the dose of the environmental agent is greater than 0 does the GRF have any role, and then it is to modify the effect of the exposure. In the terminology of Khoury, Rg (defined as the relative risk of the GRF alone) is defined as equal to 1. In a hypothetical 2×2 table (Table 2) showing odds ratios (OR) of disease as a function of both a genetic and an environmental

Table 2. Hypothetical table for type 2 gene–environment interaction

Exposure	GRF –	GRF +
–	1.0 [a]	1.0
+	2.0	5.0

[a] Reference cell.

risk factor, the reference cell (OR = 1) is that where neither gene nor exposure is present. For type 2 GEI, the cell of gene = 1 (present) and exposure = 0 (absent) also has an OR = 1. This is a fundamentally different situation with respect to epidemiological analysis from the case with those types of GEI where the GRF can, by itself, cause disease.

Dose effects in type 2 gene–environment interaction

A common approach to the investigation of genetic susceptibility towards environmental carcinogens is to use a case–control or other study design to determine the odds ratio for disease (cancer) associated with a particular genetic polymorphism. In many of these studies, environmental exposures such as cigarette smoking are taken into account and analysed for interaction by multiple regression. Many authors have inappropriately considered such exposure to be confounding (London *et al.*, 1995). For the metabolic genes, whose activity is not independent but in fact completely dependent on exposure, it is an error to analyse the effect of the gene correcting for exposure as a confounder. If a true type 2 GEI is present, then no effect of the GRF is expected in the absence of exposure. A spurious effect of the GRF will be found if data are adjusted for exposure considered as a confounder.

When the dose of environmental exposure (such as smoking) is analysed with respect to genotype of a metabolic susceptibility gene, two apparently divergent patterns are seen. The first instance could be described as a decreasing interaction with dose. This is seen when the proportion of cases with the genetic risk factor (GRF) have lower exposure doses than the proportion of cases without the GRF. The phenomenon occurs when the multiplicative effect of being GRF+ (versus GRF–) on the disease–exposure odds ratio becomes less as the exposure dose increases. If the end-point is not cancer but, for example, some marker of exposure such as adducts, then subjects with the polymorphism tend to have higher relative levels of adducts at lower doses of exposure, while at high exposure no difference in end-point will be observed between those with and without the GRF. While, at first, such a result may appear to be counterintuitive, it can be explained by the increased sensitivity (due to higher genetic susceptibility) of individuals with the GRF to lower levels of exposure. The decreasing interaction with dose has been observed for smoking with *CYP1A1* and lung cancer (Nakachi *et al.*, 1993; Taioli *et al.*, 1995a), for *NAT2* and haemoglobin adducts (Vineis *et al.*, 1994), and for *CYP1A2* and adducts (Landi *et al.*, personal communication).

A direct dose effect of the gene is observed when cases with the GRF have environmental exposure doses that are higher than cases without the GRF; the higher the dose, the greater is the effect of having the GRF on any other end-point (such as disease, adducts, etc.) This can be explained simply by stating that the greater gene–environment interaction seen at higher environmental doses increases the probability of an individual becoming a case beyond that seen due to exposure alone. This direct effect has been seen with *GSTM1* and lung cancer (Hirvonen *et al.*, 1993; Kihara *et al.*, 1994). However, a study of *GSTM1* and asbestosis (Smith *et al.*, 1994b) showed a decreasing interaction with dose, suggesting that this phenomenon is not simply gene-specific, but must be related to the mechanism of action of the gene product leading to the end-point being measured. In Fig. 2, we have plotted the frequency of cases (for cancer as the end-point) or of subjects with adduct values above the median (for adduct levels as the end-point) who are positive for a particular GRF as a function of exposure level from six separate studies of metabolic gene polymorphisms as biomarkers of susceptibility. The curves slope upwards in two cases (Hirvonen *et al.*, 1993; Kihara *et al.*, 1994) which both report on *GSTM1* and lung cancer. For the other four examples, two using *CYP1A1* and lung cancer (Nakachi *et al.*, 1993; Taioli *et al.*,

1995a), one for *NAT2* and haemoglobin adduct (Vineis *et al.*, 1994), and one for *GST* and asbestosis (Smith *et al.*, 1994b), the curves slope downwards as a function of dose.

GEI dose effect in relation to carcinogenesis dose–response models

Analysis of these two apparently divergent scenarios reveals that they may not be contradictory. One possible explanation is related to the position on the carcinogenesis dose–response curve at which the GRF is functioning. While the actual shape of the carcinogenicity dose–response curve is unknown for any human cancer, and is the subject of considerable research and discussion, certain general principles may be agreed on. At low doses (leaving aside the question of a threshold by assuming that we are speaking of a region beyond a putative threshold), we may say that increasing dose leads to increasing response. Whether the shape of this part of the curve is linear or follows some other function is not important to this discussion. At very high doses, the response must begin to level off, because the incidence approaches either 100% or some other biologically determined maximal value. As one approaches this maximal response, increasing the dose has less of an effect on the disease probability than is seen at lower doses. In all cases of carcinogenesis or any other toxicological end-point, a saturating dose must always exist at which no further effect can be seen at higher doses. We can assume that the effect of a genetic susceptibility factor is to increase the carcinogenic response at any particular dose, for example by causing increased enzymatic activity or by altering the metabolic profile of an agent. An important assumption here is that the GRF has no effect on the maximal response, but instead shifts the dose–response curve to the left. At lower dose levels, when the risk of the outcome is a small fraction of the maximal value, the curves will appear almost linear, and the GRF leads to an increase in the slope. Therefore, in this region, subjects with the GRF should respond more to higher doses of environmental agents than subjects without the GRF. However, if one is working at dose levels that produce disease (or other end-point) risks closer to the maximum level, than the converse will be true. In this region, although the overall response in GRF-positive individuals is higher than in GRF negatives, there is less effect of increasing dose in the positives than in the negatives. Therefore, subjects with the GRF may exhibit disease (or high adduct levels, etc.) at lower doses of environmental agent than those without the GRF.

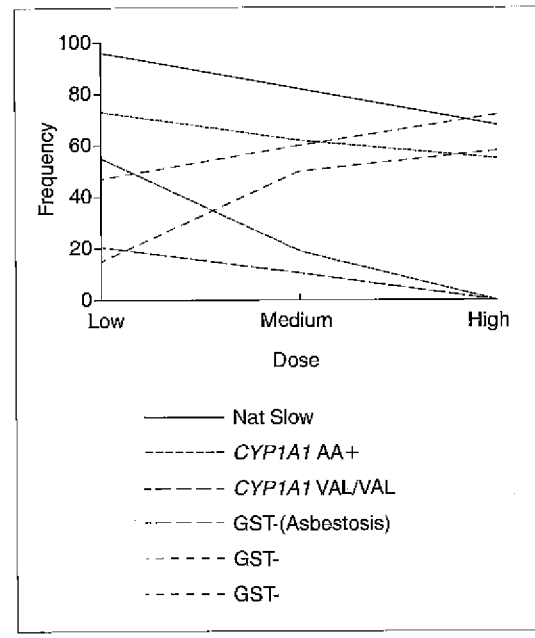

Figure 2. GRFs and smoking doses.

According to this model, genes such as *CYP1A1* or *NAT2* show an inverse dose effect with respect to lung cancer or haemoglobin adducts, because the doses of carcinogen present in cigarette smoke are so high. The effect of *GSTM1* on lung cancer with smoking is not as clear since the dose is also high. One possibility is that the substrate for *GST* is far from the saturable level even when the external dose is very high, as in the case of smokers. According to this model, a direct dose effect would be seen for *CYP1A1* polymorphisms with cancer related to low exposure doses of carcinogenic hydrocarbons, and an inverse dose effect might be observed for *GSTM1* (as is apparently the case for asbestosis) for the appropriate mechanistic pathway.

It is important to understand clearly the meaning of the inverse dose effect in type 2 GEI to avoid a confusing message. For a GRF that exhibits an inverse dose effect associated with smoking and lung cancer, the following may be said: individuals

who are positive and smoke only a few cigarettes a day are at relatively greater risk, *compared with GRF negatives,* than if they smoke two packs a day. This does not mean that such people should smoke more, because their risk of cancer still increases with dose compared with non-smokers. In fact, an inverse dose effect implies that for GRF-positive individuals, even a low smoking dose is highly risky; people carrying the polymorphism are at higher risk of cancer in comparison to the general population exposed to smoke. Only complete smoking cessation, as well as the avoidance of other relevant exposures, can lead to cancer prevention in the susceptible group.

An approach for the analysis of type 2 GEI

In order to proceed further with analysis of type 2 gene–environment interactions for metabolic genes, it is necessary to use analytical statistical tools on case–control and other study designs (Hwang *et al.*, 1994). We propose an approach to the analysis of such studies that may have the ability to detect the specific form of dose effect in a quantitative manner. The common way to describe interactions between the effects of an environmental agent and a genetic risk factor is to use both terms in a multiple regression model, and to also include a term that multiplies the GRF by the environmental agent. The coefficient of this interactive term then determines whether interaction is present, as illustrated in equation (1):

$$G(Y) = a + b_1X_1 + b_2X_2 + b_3X_1X_2 \qquad (1)$$

where Y is the odds of disease, X_1 is the environmental exposure, and X_2 is the GRF. The coefficients b_1, b_2 and b_3 are determined by regression analysis using an appropriate computer program. If we accept by definition that in the absence of environmental exposure the presence of the GRF by itself has no effect on disease outcome (the definition of a type 2 interaction), then b_2 in the regression model of equation (1) is defined as equal to zero. Thus equation (1) becomes:

$$G(Y) = a + b_1X_1 + b_3X_1X_2 \qquad (2)$$

which can be written in a different way:

$$G(Y) = a + (b_1 + b_3X_2)X_1 = a + b^*X_1 \qquad (3)$$

This expression corresponds to the assuption that the risk of disease is due only to the action of the environmental exposure, and the only effect of the GRF X_2 is to modify the coefficient of the exposure term. Now we can say that:

$$b^* = b_1 + b_3X_2 = b_1(1+ \alpha X_2), \qquad (4)$$

where $\alpha = b_3/b_1$.

For example, let us assume that the effect of GRF such as the ILE to VAL polymorphism in the *CYP1A1* gene is to increase the enzymatic activity of the gene product, as has been shown (Cosma *et al.*, 1993a; Crofts *et al.*, 1994; Landi *et al.*, 1994; Taioli *et al.*, 1995b). The result of having this GRF is an increased level of metabolism, presumably leading to an increased concentration of the ultimate carcinogen, given a particular exposure dose. While this scenario may represent an oversimplification, it can be seen that the effect of the GRF is to quantitatively modify the effect of the exposure term. This would be reflected in a value for the term α that is greater than 0. Note that if α is negative, the genetic factor would be protective. If the GRF has no effect on the exposure (for example, if the exposure is to an agent that is not a substrate for the gene product), then $\alpha = 0$. If the GRF is absent, then $X_2 = 0$. In either case, the risk is a function of exposure only, with no contribution from the gene.

Now, we can rewrite the regression model of equation (2) if there are data for the effects of multiple (n) levels of exposure (doses) as:

$$G(Y) = a + be_1E_1 + be_2E_2 + ...+ be_nEn + b(eg)_1GE_1 \; b(eg)_2GE_2 +...+ b(eg)_nGEn \qquad (5)$$

where $be_i = b_1$ for exposure level E_1, (termed X_1 in equation 2), $b(eg)_1 = b_3$ for exposure level E_1, and G stands for the GRF (X_2), etc. Using this notation, E_i stands for dose, and the term for bg (or b_2 in equation 1) is defined as 0 and does not appear. For each dose level, from equation (4), the term α is equal to:

$$\alpha_i = b(eg)_i/be_i \qquad (6)$$

If values of α are plotted against dose, several outcomes are possible. If the slope of this plot is positive, then the gene–environment interaction

Table 3. Odds ratios for lung cancer as a function of GSTM1 genotype and level of smoking exposure

		Exposure			
Genotype	**GRF**	**None**	**Low**	**Medium**	**High**
GSTM1+	–	ND	1.0 [a]	1.71	1.71
GSTM1– (null allele)	+	ND	1.03	3.4	5.04
CYP1A1 ILE	–	1.0 [a]	2.83	18.6	33.1
CYP1A1 VAL	+	0.93	22.2	40.0	40.1

[a]Reference cell.

follows a direct dose effect. If the slope is negative, then an inverse dose effect is operative. Another way to express this is to say that, if the term $\alpha_{i+1}/\alpha_i > 1$, then the dose effect is direct; if $\alpha_{i+1}/\alpha_i < 1$, then there is an inverse dose effect. If α is less than 0 at any particular dose level, then the genetic factor is protective at that level. Such a scenario, whereby a particular genetic polymorphism may be a risk factor at one level of exposure but protective at a different level, is possible given the highly complex web of interconnecting metabolic pathways that usually operate in carcinogenic mechanisms.

Clearly, we assume that α is some function of dose, but the particular function is likely to vary for every GRF and for different specific chemical exposures. It is unlikely that the limited amount of dose–response data that is usually available from case–control studies can allow for a precise definition of this function. However, this approach is useful in the characterization of genetic risk factors in terms of their dose effects, and, especially, has the advantage of avoiding confusion between an inverse dose effect and a protective effect. Two examples from the literature will be used to illustrate the method.

Examples of type 2 GEI analyses

We will use the data from Kihara *et al.* (1994) to illustrate a direct dose effect. Table 3 shows the odds ratios (ORs) for each category of smoking exposure and genotype. The first point is that the OR for the cell where the GRF is present but where there is no exposure is equal to the reference OR for the absence of gene and exposure. This marks the interaction as a type 2 interaction. We see that,

Table 4. Coefficients from regressionanalysis for gene–environment interaction

Exposure level	**Coefficient**	**Example 1 (*GSTM1*)**	**Example 2 (*CYP1A1*)**
1	be_1	0.535	1.039
2	be_2	0.535	2.92
3	be_3	–	3.5
1	$b(eg)_1$	0.66	2.06
2	$b(eg)_2$	1.05	0.76
3	$b(eg)_3$	–	0.190
1	α_1	1.233	1.98
2	α_2	1.96	0.26
3	α_3	–	0.054

as exposure level increases, the risk of disease increases, and that the increase is higher when the GRF is present for each category of exposure level. Table 4 shows the coefficients [*be*, *b*(*eg*)] obtained from the multiple logistic regression model using the SAS statistical package Genmod. Also shown in the table are the values for α_i the interaction term, which increase directly as a function of dose. Thus, for this case, there is a direct dose effect. Similar results can be obtained using the data from other sources for this gene and smoking-related lung cancer (Hirvonen *et al.*, 1993).

An example of an inverse dose effect is seen for *CYP1A1* ILE to VAL polymorphism in exon 7 of the gene, as a GRF for smoking-induced lung cancer. The data from Nakachi *et al.* (1991) are shown in Tables 3 and 4. Here, although the OR for cancer increases for both genotypes as a function of dose, the ratio between the risks for the two genotypes decreases at higher doses. The decrease of α with increasing exposures illustrates this. We have also observed an inverse dose effect for the association of the African-American-specific polymorphism in *CYP1A1* with lung adenocarcinoma in smokers (Taioli *et al.*, 1995a), and other groups have reported similar findings using adducts as an end-point for *NAT2* and for *CYP1A2* (Vineis *et al.*, 1994; Landi *et al.*, personal communication).

In some cases, this analytical approach may be used to determine whether a type 2 GEI is in fact the correct model for a particular case study. This may be done by determining if the odds ratio for the unexposed group with the GRF is close to 1. In the first example used for illustration (for *GSTM1*), this was the case. Unfortunately, however, most studies of this type use very small numbers of subjects, and the data may be of insufficient statistical power to allow for an accurate determination of this odds ratio. This is especially true for studies of lung cancer and smoking, where it is often difficult to find sufficient numbers of non-exposed (non-smokers) cases of both genotypes to have adequate power. In our second example (*CYP1A1*) there were only two non-smokers among the cases! We therefore used hypothetical data, assuming a true type 2 interaction with an odds ratio close to 1 for the $G = 1$, $E = 0$ cell. Given the difficulty in proving a type 2 interaction by the use of actual data, one alternative is to apply mechanistic knowledge regarding the mechanism of action of the gene related to the exposure to decide whether a type 2 interaction is logical within a mechanistic context.

Summary and conclusions

The study of metabolic gene polymorphisms as cancer susceptibility genotypes is likely to expand in the future, given the advances in PCR-based technology and the expected advances in knowledge of the human genome. In terms of gene–environment interactions, it is critical to develop tools with which to define precisely the role of susceptibility biomarkers in cancer causation. Some such markers may have little relevance to carcinogenic exposures, while others may be involved in complex associations that are difficult to unravel. We have put forward the argument here that type 2 GEI, for which the genotype alone, in the absence of a relevant carcinogenic exposure, plays no role in carcinogenesis, is the most important type of GEI for those genes that currently make up the major category of cancer susceptibility biomarkers. We have stressed the importance of exposure dose in the analysis of GEI for these genes, and discussed two different forms of dose effect. We have shown how these forms, the direct dose effect and the inverse dose effect, may be analysed and distinguished from each other and from other types of effects (such as protection) by using regression analysis. We have suggested that the apparent divergence between these two types of dose effect may be due to the mechanistic position of the GRF on a saturable dose–response curve for cancer induction.

We have not addressed certain issues related to regression analysis of interaction. In general, questions of statistical inference using this analysis have not been addressed here, but of course such questions are important in testing hypotheses and require careful investigation. The question of whether multiplicative versus additive models of interaction should be used (Brown & Chu, 1989) has also not been addressed here. Although a multiplicative model was used to obtain the values shown in Table 4, we have found (not shown) that the use of an additive model has an effect only on the magnitude and standard errors of the coefficients and α-values, but not on the direction of the dose-dependent effect of the gene.

In conclusion, we believe that studies investigating the possibility of a metabolic gene polymorphism as a GRF should always include as much

information as possible on relevant exposure levels. Even if quantitative measures are unobtainable, questionnaire data giving approximate levels (such as never, low or high) of exposure information can be of use in the analysis of the dose effect. Comparison of α-values between studies and meta-analysis of α for specific gene–exposure combinations may also prove valuable in the future.

It has been pointed out by several authors in the field of biomarkers in epidemiology that the detection of cancer genetic susceptibility of this type has profound positive public health implications for cancer prevention. Detailed study of the interactions of these genes with environmental carcinogens promises to allow epidemiologists to consider humans as collections of individuals of varying sensitivity and responsiveness. While the entire issue of genetic susceptibility differences among people has important ethical, legal and political issues, we believe that increased knowledge in this area (such as the specific form of dose effect discussed here) will provide benefits in helping to resolve these issues.

References

Alexandrie, A.K., Sundberg, M.I., Seidegard, J., Tornling, G. & Rannug, A. (1994) Genetic susceptibility to lung cancer with special emphasis on CYP1A1 and GSTM1: a study on host factors in relation to age at onset, gender and histological cancer types. *Carcinogenesis*, 15, 1785–1790

Brennan, P. & Silman, A.J. (1994) An investigation of gene-environment interaction in the etiology of rheumatoid arthritis. *Am. J. Epidemiol.*, 140, 453–460

Brown, C.C. & Chu, K.C. (1989) Additive and multiplicative models and multistage carcinogenesis theory. *Risk Anal.*, 9, 99–105

Caporaso, N. (1996) Genetic susceptibility and the common cancers. *Biomarkers*, 1, 174–177

Catteau, A., Bechtel, Y.C., Poisson, N., Bechtel, P.R. & Bonaiti-Pellie, C. (1995) A population and family study of CYP1A2 using caffeine urinary metabolites. *Eur. J. Clin. Pharmacol.*, 47. 423–430

Conney, A.H., Miller, E.C. & Miller, J.A. (1956) The metabolism of methylated aminoazo dyes. V. Evidence for induction of enzyme synthesis in the rat by 3-methylcholanthrene. *Cancer Res.*, 16, 450–460

Conway, K., Edmiston, S., Fried, D.B., Hulka, B.S., Garrett, P.A. & Liu, E. (1995) Ha-ras rare alleles in breast cancer susceptibility. *Breast Cancer Res. Treat.*, 35, 97–104

Cosma, G. Crofts, F., Currie, D. Toniolo, P. & Garte, S.J. (1993a) The relationship between genotype and function of the human CYP1A1 gene. *J. Tox. Environ. Health*, 40, 315–322

Cosma, G., Crofts, F., Currie, D., Wirgin, I., Toniolo, P. & Garte, S.J. (1993b) Racial differences in restriction fragment length polymorphisms and messenger RNA inducibility of the human CYP1A1 gene. *Cancer Epidemiol. Biomarkers Prev.*, 2, 53–57

Crofts, F., Cosma, G.N., Taioli, E., Currie, D.C., Toniolo, P.T. & Garte, S.J. (1993) A novel CYP1A1 gene polymorphism in African-Americans. *Carcinogenesis*, 14, 1729–1731

Crofts, F., Taioli, E., Trachman, J. Cosma, G.N., Currie, D., Toniolo, P. & Garte, S.J. (1994) Functional significance of different human CYP1A1 genotypes. *Carcinogenesis*, 15, 2961–2963

Daly, A.K., Cholerton, S., Armstrong, M. & Idle, J.R. (1994) Genotyping for polymorphisms in xenobiotic metabolism as a predictor of disease susceptibility. *Environ. Health Perspect.*, 102 (Suppl. 9P), 55–61

Doll, R. & Peto, R. (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl Cancer Inst.*, 66, 1191–1308

Frebourg, T. & Friend, S.H. (1992) Cancer risks from germline p53 mutations. *J. Clin. Invest.*, 90, 1637–1641

Friend, S.H., Dryja, T.P. & Weinberg, R.A. (1988) Oncogenes and tumor-suppressing genes. *N. Engl. J. Med.*, 318, 618–622

Garte, S.J. (1992) Environmental carcinogenesis. In: Rom, W, ed., *Environmental and Occupational Medicine*. Boston, Little Brown, pp. 105–124

Garte, S.J. & Kneip, T.J. (1988) Metabolism. In: Kneip, T.J. & Crable, J.V., eds, *Methods for Biological Monitoring*. Washington, American Public Health Association, pp. 15–26

Haldane, J.B.S. (1946) The interaction of nature and nurture. *Ann. Eugenics*, 13, 197–205

Harris, C.C. (1989) Interindividual variation among humans in carcinogen metabolism, DNA adduct formation and DNA repair. *Carcinogenesis*, 10, 1563–1566

Hayashi, S.-I., Watanabe, J., Nakachi, K. & Kawajiri, K. (1991) Genetic linkage of lung-cancer-associated Msp1 polymorphims with amino acid replacement in the heme binding region of the human cytochrome P4501A1 gene. *J. Biochem.*, 110, 407–411

Hayden, M.R, Liu, M.S & Ma, Y. (1994) Gene environment interaction and plasma triglyceride levels: the crucial role of lipoprotein lipase. *Clin. Genet.*, 46, 15–18

Hegele, R.A. (1992) Gene-environment interactions in

atherosclerosis. *Mol. Cell. Biochem.*, 113, 177–186

Higginson, J. (1980) Importance of environmental and occupational factors in cancer. *J. Toxicol. Environ. Health*, 6, 941–952

Hirayama, T. (1989) Family history in childhood malignancies with special reference to genetic-environmental interactions. In: Lynch, H.T. & Hirayama, T., eds, *Genetic Epidemiology of Cancer*. Boca Raton, FL, CRC Press, pp. 1111–1117

Hirvonen, A. (1995) Genetic factors in individual responses to environmental exposures. *J. Occup. Environ. Med.* 37, 37–43

Hirvonen, A., Pursianen, K.H., Karjalainen, A., Anttila, S. & Vainio, H. (1992) Point mutational MspI and Ile-Val polymorphism closely linked in the CYP1A1 gene: lack of association with susceptibility to lung cancer in a Finnish study population. *Cancer Epidemiol. Biomarkers Prev.*, 1, 485–489

Hirvonen, A., Husgafvel-Pursianen, K., Anttila, S. & Vainio, H. (1993) The GSTM1 null genotype as a potetial risk midifier for squamous cell carcinoma of the lung. *Carcinogenesis*, 14, 1479–1481

Hollstein, M.C., Wild, C.P., Bleicher, F., Chutimataewin, S., Harris, C.C., Srivatanakul, P. & Montesano, R. (1993) p53 mutations and aflatoxin B.1 exposure in hepatocellular carcinoma patients from Thailand. *Int. J. Cancer*, 53, 51–55

Hulka, B.S. (1991) Epidemiological studies using biological markers: issues for epidemiologists. *Cancer Epidemiol. Biomarkers Prev.*, 1, 13–19

Hunt, S.C., Hasstedt, S.J., Wu, L.L. & Williams, R.R. (1993) A gene-environment interaction between inferred kallikrein genotype and potassium. *Hypertension*, 22, 161–168

Hwang, S.J., Beaty, T.H., Liang, K.Y., Coresh, J. & Khoury, M.J. (1994) Minimum sample size estimation to detect gene-environment interaction in case-control designs. *Am. J. Epidemiol.*, 140, 1029–1037

Hwang, S.J., Beaty, T.H., Panny, S.R., Street, N.A., Joseph, J.M., Gordon, S., McIntosh, I. & Francomano, C.A. (1995) Association study of transforming growth factor alpha (TGF alpha) TaqI polymorphism and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. *Am. J. Epidemiol.*, 141, 629–636

Ichiba, M., Hagmar, L., Rannug, A., Hogstedt, B., Alexandrie, A.K., Carstensen, U. & Hemminki, K. (1994) Aromatic DNA adducts, micronuclei and genetic polymorphism for CYP1A1 and GST1 in chimney sweeps. *Carcinogenesis*, 15, 1347–1352

Idle, J.R., Armstrong, M., Boddy, A.V., Boustead, C., Cholerton, S., Cooper, J., Daly, A.K., Ellis, J., Gregory, W., Hadidi, H. (1992) The pharmacogenetics of chemical carcinogenesis. *Pharmacogenetics*, 2, 246–258

Kato, S., Shields, P.G., Caporaso, N.E., Sugimura, H., Trivers, G.E., Tucker, M.A., Trump, B.F., Weston, A. & Harris, C.C. (1994) Analysis of cytochrome P4502E1 genetic polymorphisms in relation to human lung cancer. *Cancer Epidemiol. Biomarkers Prev.* 3, 515–518

Kawajiri, K., Nakachi, K., Imai, K., Yoshii, A., Shinoda, N. & Watanabe, J. (1990) Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome p4501A1 gene. *FEBS Lett.*, 263, 131–133

Khoury, M.J., Adams, M.J. & Flanders, W.D. (1988) An epidemiologic approach to ecogenetics. *Am. J. Hum. Genet.*, 42, 89–95

Khoury, M.J., Beaty, T.H. & Cohen, B.H. (1993) *Fundamentals of Genetic Epidemiology*. New York, Oxford University Press

Khoury, M.J. & James, L.M. (1993) Population and familial relative risks of disease associated with environmental factors in the presence of gene-environment interaction. *Am. J. Epidemiol.*, 137, 1241–1250

Khoury, M.J., Beaty, T.H. & Hwang, S.J. (1995) Detection of genotype-environment interaction in case-control studies of birth defects: how big a sample size? *Teratology*, 51, 336

Khoury, M.J. & Wagener, D.K. (1995) Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. *Am. J. Hum. Genet.*, 56, 835–844

Kihara, M., Kihara, M. & Noda, K. (1994) Lung cancer risk of GSTM1 null genotype is dependent on the extent of tobacco smoke exposure. *Carcinogenesis*, 15, 415–418

Landi, M.T., Bertazzi, P.A., Shields, P.G., Clark, G., Lucier, G.W., Garte, S.J., Cosma, G. & Caporaso, N.E. (1994) Restriction fragment length polymorphism, mRNA expression and enzymatic product of CYP1A1 gene in humans. *Pharmacogenetics*, 4, 242–246

Lane A. Cruickshank J.K. Mitchell J. Henderson A. Humphries S. Green, F. (1992) Genetic and environmental determinants of factor VII coagulant activity in ethnic groups at differing risk of coronary heart disease. *Atherosclerosis*, 94, 43–50

Lehmann, A.R., Hoeijmakers, J.H., van Zeeland, A.A., Backendorf, C.M., Bridges, B.A., Collins, A., Fuchs, R.P., Margison, G.P., Montesano, R. & Moustacchi, E. (1992) Workshop on DNA repair. *Mutation Res.* 273, 1–28

London, S.J., Daly, A.K., Fairbrother, K.S., Holmes, C., Carpenter, C.L., Navidi, W.C. & Idle, J.R. (1995) Lung cancer risk in African-Americans in relation to a race-specific CYP1A1 polymorphism. *Cancer Res.*, 55, 6035–6037

Marquis, J.K. & Siek, G.C. (1988) Sensitive populations and risk assessment in environmental policy-making. In: Saxena, J., ed., *Hazard assessment of chemicals* (Vol. 6). Washington, DC, Hemisphere Publishing

Meyer (1994) Polymorphisms of human acetyltransferases. *Environ. Health Perspect.*, 102 (Suppl. 6P), 213–216

Nakachi, K., Imai, K., Hayashi, S., Watanabe, J. & Kawajiri, K. (1991) Genetic susceptibility to squamous cell carcinoma of the lung in relation to cigarette smoking dose. *Cancer Res.*, 51, 5177–5181

Nakachi, K., Imai, K., Hayashi, S. & Kawajiri, K. (1993) Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. *Cancer Res.*, 53, 2994–2999

Nakajima, T., Elovaara, E., Anttila, S., Hirvonen, A., Camus, A.M., Hayes, J.D., Ketterer, B. & Vainio, H. (1995) Expression and polymorphism of glutathione S-transferase in human lungs: risk factors in smoking-related lung cancer. *Carcinogenesis*, 16, 707–711

Ottman, R. (1990) An epidemiologic approach to gene-environment interaction. *Genet. Epidemiol.*, 7, 177–185

Ottman, R. (1994) Epidemiologic analysis of gene-environment interaction in twins. *Genet. Epidemiol.*, 11, 75–86

Ottman, R. (1995) Gene-environment interaction and public health [editorial; comment]. *Am. J. Hum. Genet.*, 56, 821–823

Perera, F., Mayer, J., Santella, R.M., Brenner, D., Jeffrey, A., Latriano, L., Smith, S., Warburton, D., Young, T.L., Tsai, W.Y., Hemminki, K. & Brandt-Rauf, P. (1991) Biologic markers in risk assessment for environmental carcinogens. *Environ. Health Perspect.*, 90, 247–254

Petersen, D.D., McKinney, C.E., Ikeya, K., Smith, H.H., Bale, A.E., McBride, O.W. & D. W. Nebert. (1991) Human CYP1A1 gene: cosegregation of the enzyme inducibility phenotype and an RFLP. *Am. J. Hum. Genet.*, 48, 720–725

Raunio, H., Husgafvel-Pursiainen, K., Anttila, S., Hietanen, E., Hirvonen, A. & Pelkonen, O. (1995) Diagnosis of polymorphisms in carcinogen-activating and inactivating enzymes and cancer susceptibility—a review. *Gene*, 159, 113–121

Rothman, N. (1995) Genetic susceptibility biomarkers in studies of occupational and environmental cancer: methodologic issues. *Toxicol. Lett.* 77, 221–225

Samson, L. (1992) The suicidal DNA repair methyltransferases of microbes. *Mol. Microbiol.*, 6, 825–831

Savitsky K. Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., & Sfez, S. (1995) A single ataxia telangiectasia gene with a product similar to PI-3kinase. *Science*, 268, 1749–1753

Sellers, T.A., Potte, J.D., Bailey-Wilson, J.E., Rich, S.S., Rothschild, H. & Elston, R.C. (1992) Lung cancer detection and prevention: evidence for an interaction between smoking and genetic predisposition. *Cancer Res.*, 52, 2694S–2697S

Shields, P.G., Caporaso, N.E., Falk, R.T., Sugimura, H., Trivers, G.E., Trump, B.F., Hoover, R.N., Weston, A. & Harris, C.C. (1993) Lung cancer, race, and a CYP1A1 genetic polymorphism. *Cancer Epidemiol. Biomarkers Prev.*, 2, 481–85

Sinha, R., Rothman, N., Brown, E.D., Mark, S.D., Hoover, R.N., Caporaso, N.E., Levander, O.A., Knize, M.G., Lang, N.P. & Kadlubar, F.F. (1994) Pan-fried meat containing high levels of heterocyclic aromatic amines but low levels of polycyclic aromatic hydrocarbons induces cytochrome P.4501A2 activity in humans. *Cancer Res.*, 54, 6154–6159

Smith, M.L., Chen, I.T., Zhan, Q., Bae, I., Chen, C.Y., Gilmer, T.M., Kastan, M.B., O'Connor, P.M. & Fornace A.J., Jr (1994a) Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science*, 266, 1376–1380

Smith, C.M., Kelsey, K.T., Wiencke, J.K., Leyden, K., Levin, S. & Christiani, D.C. (1994b) Inherited glutathione-S-transferase defiency as risk factor for pulmonary asbestosis. *Cancer Epidemiol. Biomarkers Prev.*, 3, 471–477

Stephens, E.A., Taylor, J.A., Kaplan, N., Yang, C.H., Hsieh, L.L., Lucier, G.W. & Bell, D.A. (1994) Ethnic variation in the CYP2E1 gene: polymorphism analysis of 695 African-Americans, European-Americans and Taiwanese. *Pharmacogenetics*, 4, 185–192

Sugimura, H., Suzuki, I., Hamada, G.S., Iwase, T., Takahashi, T., Nagura, K., Iwata, H., Watanabe, S., Kino, I. & Tsugane, S. (1994) Cytochrome P4501A1 genotype in lung cancer patients and controls in Rio de Janeiro Brazil. *Cancer Epidemiol. Biomarkers Prev.*, 3, 145–148

Taioli, E. Crofts, F. Demopoulos, R., Trachman, J. Toniolo, P. & Garte, S.J. (1995a) An African American specific CYP1A1 polymorphism is associated with adenocarcinoma of the lung. *Cancer Res.*, 55, 472–473

Taioli, E. Crofts, F., Trachman, J., Bayo, S., Bradlow, L., Toniolo, P. & Garte, S.J. (1995b) Racial differences in CYP1a1 genotype and function. *Toxicol. Lett.*, 77, 357–362

Taioli, E., Trachman, J., Chen, X., Toniolo, P. & Garte, S.J. (1995c) CYP1A1 RFLP is associated with breast cancer in African American women. *Cancer Res.*, 55, 3757–3758

Tefre, T., Ryberg, D., Haugen, A., Nebert, D.W., Skaug, V., Brogger, A. &, Borresen, A.L. (1991) Human CYP1A1 gene: lack of association between the Msp1 restriction fragment

length polymorphism and incidence of lung cancer in a Norwegian population. *Pharmacogenetics*, 1, 20–27

Tiret, L., Abel, L. & Rakotovao, R. (1993) Effect of ignoring genotype-environment interaction on segregation analysis of quantitative traits. *Genet. Epidemiol.*, 10, 581–586

Vahakangas, K.H., Samet, J.M., Metcalf, R.A., Welsh, J.A., Bennett, W.P., Lane, D.P. & Harris, C.C. (1992) Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. *Lancet*, 339, 576–580

Vineis, P. (1995) Use of biomarkers in epidemiology. The example of metabolic susceptibility to cancer. *Toxicol. Lett.*, 77, 163–168

Vineis, P., Bartsch, H., Caporaso, N., Harrington, A., Kadlubar, F.F., Landi, M.T., Malaveille, C., Shields, P.G., Skipper, P., Talaska, G. & Tannenbaum, S.R. (1994) Genetically based N-acetyltransferase metabolic polymorphism and low level environmental exposure to carcinogens. *Nature*, 369, 154–156

Watanabe, J., Yang, J.P., Eguchi, H., Hayashi, S., Imai, K., Nakachi, K. & Kawajiri, K. (1995) An Rsa I polymorphism in the CYP2E1 gene does not affect lung cancer risk in a Japanese population. *Jpn J. Cancer Res.*, 86, 245–248

Yu, M.C., Skipper, P.L., Taghizadeh, K., Tannenbaum, S.R., Chan, K.K., Hendeson, B.E. & Ross, R.K. (1994) Acetylator phenotype, aminobiphenyl-hemoglobin adduct levels, and bladder cancer risk in white, black, and Asian men in Los Angeles California. *J. Natl Cancer Inst.*, 86, 712–716

Corresponding author

S. Garte
Nelson Institute of Environmental Medicine and
Kaplan Cancer Center,
New York University Medical Center,
New York, NY, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Using and interpreting surrogate end-points in cancer research

A. Schatzkin, L.S. Freedman, J. Dorgan, L. McShane, M.H. Schiffman and S.M. Dawsey

Researchers have proposed a broad range of molecular, cellular and histological markers as surrogate end-points for cancer (SECs). The effect of an intervention on a 'valid' SEC is concordant with its effect on cancer incidence. The validity of a potential SEC is determined primarily by the extent to which the marker is a necessary event on the causal pathway to cancer. Colorectal adenomatous polyp formation is an example of a reasonably valid SEC because these lesions are obligate precursors of most large bowel malignancies. However, the existence of a plausible major alternative causal pathway—one bypassing the potential SEC—weakens inferences from that marker to cancer. Moreover, unless the pathway to cancer operates nearly exclusively through the SEC, an SEC that is valid for one intervention or exposure may not be valid for another. Metabolic, ecological, observational epidemiological and intervention studies may yield data that are useful in revealing these causal interrelations of intervention (exposure), SEC and cancer. Empirical studies of three questions are pertinent: (1) What is the relation of the SEC to cancer? (2) What is the relation of the intervention (exposure) to the SEC? (3) To what extent does the SEC mediate the relation between the intervention (exposure) and cancer? Data on SEC measurement error are important in ascertaining the extent to which marker results have been attenuated by such error. It is essential to carry out these studies to evaluate potential SECs (such as epithelial cell hyperproliferation) with plausible major alternative pathways to cancer. At the present time, definitive evidence on etiology and prevention will emerge only from studies with cancer end-points or SECs that are, by and large, necessary steps on the causal pathway to malignant disease.

Because the diagnosis of cancer is a relatively rare event, clinical trials or observational epidemiological studies with incident cancer end-points have to be very large, lengthy and expensive. Studies using surrogate end-points of malignant disease can be smaller, shorter and cheaper than studies with incident cancer end-points. It is not surprising, then, that cancer researchers have long been interested in using these markers.

Definition of a surrogate end-point marker for cancer

We define a surrogate end-point marker for cancer (SEC) as follows: a surrogate for incident cancer yields a valid test of the null hypothesis of no association between treatment and incident cancer. In other words, the effect of an intervention on the SEC is concordant with its effect on cancer incidence, or, for observational epidemiological studies, the association of an exposure with the SEC is concordant with its association with cancer incidence. If, for example, a large change in the SEC means a large change in cancer incidence, then a small change in the SEC would mean a small change in cancer incidence. If the SEC meets these conditions, it can be considered a 'valid' surrogate for that cancer. These conditions follow from the criteria proposed by Prentice (1989).

Evaluating SEC validity: logical considerations

The validity of a potential SEC depends on the extent to which it is a necessary step in carcinogenesis. The simplest causal pathway involving a potential SEC is shown in Fig. 1. E1 represents an environmental or host factor. A change in E1 would alter SEC positivity and thereby modify the incidence of cancer. The SEC is a valid surrogate for cancer.

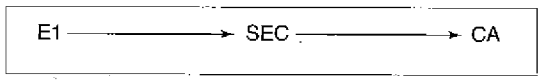

Figure 1. Pathway to cancer (CA) with single exposure (E1) and single marker (SEC).

Figure 2 depicts a more realistic picture of carcinogenesis. In this figure, E1 influences cancer through two alternative pathways, one through the potential SEC, the other through Marker 2. (We assume, to simplify the discussion, that only one intermediate marker, SEC, resides on the E1–SEC–cancer pathway. To the extent that E1 works through the alternative Marker 2 pathway, we cannot be certain that SEC is a valid surrogate for studies involving E1. This is because E1 may affect Marker 2 in a way that offsets its influence on SEC; the final effect on cancer is unknown. If, for example, E1 reduces SEC positivity but increases Marker 2 positivity, E1 could increase cancer incidence.

Potential SECs illustrating these logical considerations

Consider adenomatous polyps, an increasingly popular surrogate for colorectal cancer. In Fig. 3 (pathway a), an event X is necessary for an adenomatous polyp to progress to colorectal cancer. Therefore, two types of polypoid adenomas exist—those without X (innocent adenomas not progressing to cancer) and those with X ('bad' adenomas progressing to cancer). Both types are observable but indistinguishable through a colonoscope. Furthermore, there exist flat areas of dysplasia with X (not observable through the colonoscope) that also progress to cancer (Hill, 1991).

Suppose we have an intervention (a low-fat eating plan, for example) which reduces E1 (some faecal constituent) and thereby diminishes the pool of adenomas susceptible to the relatively rare X-events. This intervention thus reduces the number of bad adenomas as a necessary consequence of reducing the number of innocent adenomas. The intervention therefore lowers the incidence of colorectal cancer.

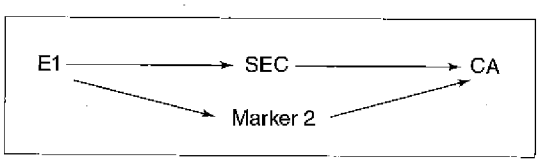

Figure 2. Single exposure (E1) with pathways through two alternative markers (SEC and Marker 2) to cancer (CA).

The existence of the flat dysplasia pathway complicates things. Our intervention has no effect on pathway (b), even though it reduces observed adenomas via pathway (a). To the extent that pathway (b) contributes to colorectal carcinogenesis, adenoma development (as detected through colonoscopy) may not be a valid SEC for cancer. As pathway (a) becomes the less common of the two routes to cancer, an investigator observing fewer adenomas developing among intervention participants could conclude that the intervention reduces colorectal cancer incidence, when in fact the intervention might have a quite different effect on cancer occurrence through pathway (b). A large body of evidence, however, suggests that most colorectal cancers do develop through pathway (a), the adenoma–carcinoma sequence (Muto *et al.*, 1975). Therefore, an intervention reducing adenomatous polyp recurrence would be likely to reduce colorectal cancer incidence. Adenoma recurrence is a reasonably valid SEC.

HPV infection in cervical cancer appears to be analogous to the adenoma–colorectal cancer example. The overwhelmingly large proportion of cervical cancer requires prior HPV infection (Schiffman, 1992). The relatively rare X-event is whatever (still unknown) immunological deficit leads to persistent HPV infection. HPV persistence results in inactivation, by the E6 and E7 proteins of the HPV genome, of *p53* and *pRb* tumour suppressor genes, leading in turn to increasingly severe intra-epithelial neoplasia and eventually cancer. It is currently thought, however, that a small proportion of cervical cancer can arise as a result of tumour suppressor gene product inactivation occurring, by mutation, in the absence of HPV infection. Because most cervical cancer does occur through HPV infection, an intervention that eliminates or reduces HPV infection would probably decrease cervical cancer incidence.

Inferences to cancer from other potential SECs, however, are considerably more problematic. Figure 4 depicts plausible causal pathways involving colorectal epithelial cell proliferation (Baron *et al.*, 1995) (the adenoma step depicted in Fig. 4 is simplified here). E1 is again an exposure amenable

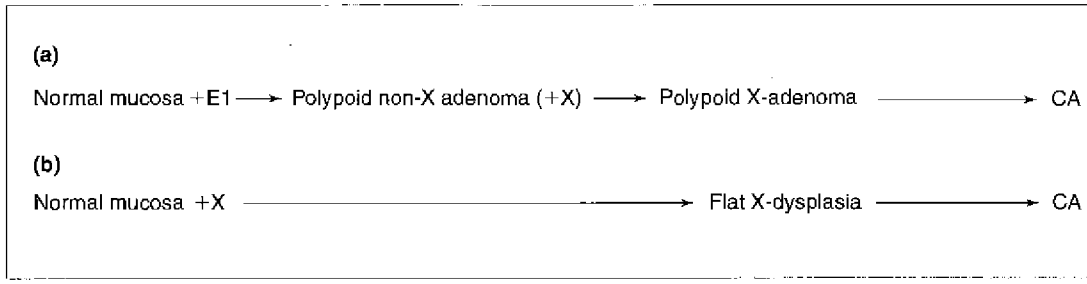

Figure 3. Alternative pathways, with (a) and without (b) an adenomatous polyp step, to colorectal cancer (CA). E1 is exposure; X is an event necessary for development of cancer.

to modification. To the extent that other cellular or molecular events, such as diminished apoptosis or altered cellular adhesion factors, constitute an important causal pathway from E1 to cancer, we cannot be sure that E1's relation to these other events does not offset E1's effect on cancer through proliferation. Cell proliferation is a problematic SEC because the relative importance of the alternative pathways (through events other than proliferation) is simply unknown.

Another logical consideration is as follows: a marker may not be directly on the causal pathway to cancer but may be closely linked to a component of that causal pathway such that it does make a reasonable SEC. One possible example of this are micronuclei, which have been detected in epithelial cells from oral, oesophageal, bronchial and large intestinal tissue (Garewal *et al.*, 1993). Many micronucleated cells are non-viable and therefore cannot be direct cellular precursors of a malignant tumor. The overall prevalence of micronucleated cells, though, might strongly reflect microstructural alterations in other cells that do eventually undergo malignant transformation and clonal expansion.

Exposure dependence

In Fig. 5 we return to the simple, idealized scheme from Fig. 1, but now add another exposure, E2. Both E1 and E2 in Fig. 5 work through SEC on the path to cancer. Because SEC is a necessary precursor for cancer, the validity of this SEC is exposure-independent, i.e. any other exposure, E2, that influences cancer must operate through the SEC; the SEC is valid for studies of E2 as well as those of E1.

Figure 6 adds E2 to the more complex and realistic pathway depicted in Fig. 2. In Fig. 6, the existence of a non-trivial alternative pathway (through Marker 2) means that the validity of the SEC is exposure-dependent. Even if E1 affects Marker 2 minimally (suggesting that SEC is reasonably valid for E1–cancer studies), we cannot assume that the E2–Marker 2–cancer pathway also plays a minor role in the development of cancer.

For example, a given agent might influence colorectal carcinogenesis largely through its influence on cell proliferation (Fig. 4). In that case, cell proliferation is a reasonably valid SEC for the first agent *vis-à-vis* colorectal cancer. A second agent, however, might not affect cell proliferation but might increase apoptosis sufficiently to decrease

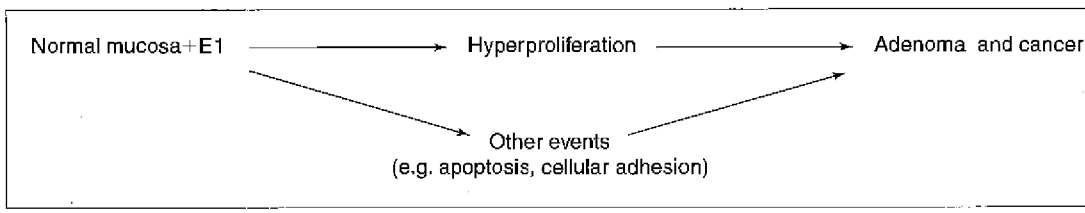

Figure 4. Alternative pathways, with and without mucosal hyperproliferation, to colorectal cancer. E1 is exposure.

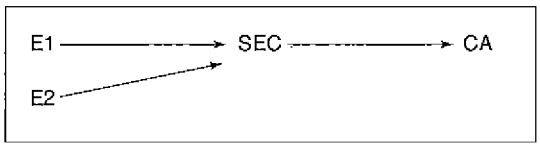

Figure 5. Each of two exposures (E1 and E2) leads through two markers (SEC and Marker 2) to cancer (CA).

cancer incidence (Bedi *et al.*, 1995). Focusing only on cell proliferation would give a falsely pessimistic impression of the second agent's efficacy in inhibiting colorectal carcinogenesis.

Investigating causal pathways involving SECs

We have argued that the causal structure underlying the relationships between exposures, potential SECs and cancer is critical in evaluating SECs. Data that are helpful in revealing this structure can emerge from investigations into three questions (Schatzkin *et al.*, 1990): (1) is the SEC associated with cancer (in particular, how large is the attributable proportion)? (2) is the intervention/exposure associated with the SEC? and (3) does the SEC mediate the relationship between the intervention/exposure and cancer ?

Traditional epidemiological parameters are useful in carrying out these investigations. For simplicity, we refer in the following discussion to SECs that are either positive or negative. However, the arguments offered here may be extended to encompass markers measured as continuous variables.

Relative risk, a measure of cancer risk in relation to SEC positivity, is defined as $[a/(a+b)]/[c/(c+d)]$ (Table 1). A relative risk of 1.0 indicates no association between SEC and cancer. The attributable proportion (AP) represents the proportion of cancer that is attributable to marker positivity: $AP = S(1 - 1/R)$, where R = relative risk and S = sensitivity, defined as $a/(a+c)$. An AP of 1.0 means that marker positivity is necessary for the development of cancer, i.e. the carcinogenic pathway must go through this positive marker.

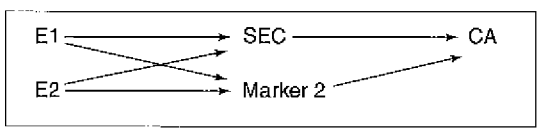

Figure 6. Each of two exposures (E1 and E2) can act through a number of pathways to affect markers (SEC and Marker 2) and lead to cancer (CA).

Table 1. 2 × 2 table showing relationship between SEC and cancer

	Cancer	
SEC	**+**	**–**
+	*a*	*b*
–	*c*	*d*

SEC–cancer

Observational epidemiological studies are important vehicles for examining this SEC–cancer question. In a recent case–control study, Schiffman *et al.* (1993) showed a markedly increased risk of severe cervical neoplasia for those with HPV infection. Toniolo *et al.* (1995), in a case–control study nested within a prospective cohort, observed a direct relationship between serum estrogens and breast cancer. Observational cohort studies may also be nested in trials. In the Polyp Prevention Trial (Schatzkin *et al.*, 1996), for example, it will be possible to relate baseline proliferation indices to subsequent adenoma recurrence. (We have referred here to studies with neoplastic cancer precursor end-points, such as cervical intra-epithelial neoplasia (CIN) and adenomas. For purposes of discussion, we consider these here as proxies for cancer, although, as we have shown, the validity of these precursor end-points is not absolute.)

Ecological studies may provide pertinent (if indirect) information, on the SEC–cancer question. Researchers have examined, for example, mean proliferative indices in groups at different risks of colorectal cancer (Lipkin *et al.*, 1984). In ecological studies, as opposed to observational studies with both marker and disease information on individuals, the link between marker and disease is indirect; one cannot be certain that those who are marker-positive are the ones with increased incidence of cancer.

The AP is of great value here in evaluating the importance of alternative pathways. In idealized Fig. 1, the AP for SEC is 1.0. In the more realistic Fig. 2, however, with at least two pathways to cancer, AP < 1.0. If AP for SEC is high, however, even if it is less than 1.0, it suggests that the alternative Marker 2 pathway plays a small role in the development of cancer. An AP substantially lower than 1.0 suggests that one or more alternative pathways is indeed operative.

Intervention/exposure–SEC

For a given SEC to be valid with respect to a given intervention, we need to demonstrate that the intervention results in a change in the SEC, or, in an observational setting, that the exposure of interest is associated with marker positivity.

We can address this question in small clinical (metabolic) studies with the putative SEC as the end-point. Examples include studies of fat (Prentice *et al.*, 1990) or alcohol (Reichman *et al.*, 1993) consumption in relation to serum hormone levels. We can also examine this question in a case–control or cohort study of, for example, the relations of reproductive risk factors to HPV infection or breast cancer risk factors to serum estrogen levels. An ecological study can examine, for example, the mean proliferative index or degree of epithelial cell DNA hypomethylation in populations with different (average) consumptions of dietary fat (Lipkin *et al.*, 1985).

Intervention/exposure–SEC–cancer

Suppose we have established that (1) the SEC is causally connected to cancer, but AP < 1.0 and the route to cancer does not proceed exclusively through the SEC; and (2) the intervention or exposure of interest is linked to the SEC. We would still like to determine the relative importance of the intervention/exposure–SEC pathway, as opposed to pathways operating through other SECs. To do this, we examine the extent to which the exposure/interventions's relation to cancer is mediated by the SEC, i.e. we address whether SEC status accounts for the observed intervention effect or exposure-associated elevation in risk. This involves integrating SEC assays into either observational epidemiological studies or clinical trials.

In a recent case–control study, for example, Schiffman *et al.* (1994) examined the extent to which HPV infection mediated the relation between number of sexual partners and cervical dysplasia. As Table 2 indicates, there was a strong direct relation between number of sexual partners and risk of cervical dysplasia. When the relation between number of sexual partners and cervical dysplasia was adjusted for the presence or absence of HPV infection, the RR for number of sexual partners dropped dramatically, suggesting that most of the relation between number of partners and dysplasia relation is attributable to HPV infection.

One can examine the mediating role of a potential SEC through stratified analyses or standard multiple regression techniques. In general, the larger the intervention effect or exposure relation, the fewer study participants are needed in a mediation analysis. Because exposure relative risks in observational epidemiological studies are often greater than the intervention effects observed in trials, mediation analyses are more likely to be successful in the observational epidemiological setting. Genetic mutations as exposures for cancer may prove to be a very fruitful source for mediation analyses of biochemical or cellular markers if they demonstrate the very high RRs that are currently predicted.

Mediation analyses may yield null results, i.e. adjusting for a potential SEC may have little influence on the relative risks for the intervention or exposure. These null findings suggest that the potential SEC does not fully mediate the relationship between intervention/exposure and cancer. Even in the face of such null results, however, when there is an alternative pathway from the exposure (E1) to cancer through a second marker,

Table 2. Cervical dysplasia odds ratio for number of sexual partners, both unadjusted and adjusted for HPV status

	Number of sexual partners				
	1	2	3–5	6–9	10+
Unadjusted	1.0	1.7	3.1[a]	4.7[a]	4.4[a]
Adjusted for HPV Status	1.0	1.0	1.1	1.5	1.6

[a] $P<0.05$.

the SEC could still reside on the causal pathway to cancer. The degree to which the E1–cancer relation is attenuated after adjustment for SEC1 will depend on the (probably unknown) extent to which the E1–cancer relation is mediated by Marker 2 as well as SEC1 (Figs 2 and 6).

Interpreting the data on SECs: statistical considerations

All markers are measured with some error. Two statistical caveats follow from this. First, a potential SEC is useful only if it can discriminate among study participants, those in an intervention and control group or those in various categories of risk factor exposure. Such discrimination is practically possible only if the interparticipant variation in the SEC values is not swamped by intra-individual variation. (Intra-individual variation derives, for example, from differences in markers obtained from different tissue areas, measured at different time points or read by multiple readers.) Statistically, this means that the intraclass correlation coefficient [ICC = interparticipant variation/(interparticipant variation + intraparticipant variation)] for interparticipant variation (the proportion of all variation attributable to between-participant differences) is reasonably large (Fleiss, 1986).

Intraparticipant variation may be reduced by taking replicate samples (multiple biopsies from different areas, multiple blood draws over time). The reduction in intraparticipant variability increases the relative contribution of the interparticipant variability, and thus the ICC.

Data on components of variance for potential SECs are very sparse. Few studies have provided data on SEC variability, particularly with respect to time-to-time variation. Notable exceptions are recent investigations attempting to determine the number of estradiol measurements necessary to discriminate reasonably among individuals (Cauley *et al.*, 1991; Toniolo *et al.*, 1994; Hankinson *et al.*, 1995). Research into rectal mucosal proliferation variability is also underway (Lyles *et al.*, 1994). We emphasize that quality control studies designed to capture information on marker variability are essential if we wish to evaluate and subsequently use a potential SEC.

Second, and more generally, even if the ratio of inter- to intraparticipant variation is acceptable, measurement error will tend to attenuate findings from each of the three types of studies discussed above. The intervention–marker and marker–cancer relations will be attenuated by error in marker measurement; the marker-adjusted intervention effect will be inflated.

Conclusion

Studies with surrogate end-points may give the right answers about the effect of an intervention or the association with an exposure. Positive results from phase 2 studies with surrogate end-points provide additional—but not incontrovertible—support for moving on to the larger, more expensive phase 3 studies with cancer end-points.

Merely being on the causal pathway to cancer does not in itself constitute surrogate validity. It is the totality of causal connections that is critical. Only when the causal pathway goes predominantly through that SEC can one reasonably make strong inferences from SEC findings to cancer. This appears to be the case for adenomas.

When there exist major alternative pathways bypassing that SEC, as in cell proliferation and many other potential SECs, inferences to cancer are problematic. This paper is in part a plea to carry out the studies—especially SEC–cancer and intervention/exposure–SEC–cancer mediation studies—necessary to evaluate those potential SECs with plausible major alternative pathways to cancer. Such studies are urgently needed if we are to know how well we can generalize from SEC results to cancer. The irony of the surrogate marker problem however, is that the large, long, expensive studies required to evaluate these problematic SECs are those the markers were designed to replace. Moreover, SEC evaluation is often intervention/exposure-dependent: results of validation studies of a marker in relation to one exposure are not necessarily transferrable to the marker in relation to another exposure/intervention. Thus, even if we were to find out from some ongoing calcium intervention studies that cell proliferation is a reasonably valid surrogate for colon cancer, we cannot be certain that proliferation indices are equally valid in aspirin or folate trials.

At present, there appears to be no substitute for carrying out large-scale epidemiological studies and clinical trials with cancer end-points or SECs that are, for the most part, necessary steps on the causal pathway to cancer.

References

Baron, J.A., Wargovich, M.J., Tosteson, T.D., Sandler, R., Haile, R., Summers, R., van Stolk, R., Rothstein, R. & Weiss, J. (1995) Epidemiological use of rectal proliferation measures. *Cancer Epidemiol. Biomarkers Prev.*, 4, 57–61

Bedi, A., Pasrich, P.J., Akhtar, A.J., Barber, J.P., Bedi, G.C., Giardiello, F.M., Zehnbauer, B.A., Hamilton, S.R. & Jones, R.J. (1995) Inhibition of apoptosis during development of colorectal cancer. *Cancer Res.*, 55, 1811–1816

Cauley, J.A., Gutai, J.P., Kuller, L.H. & Powell, J.G. (1991) Reliability and interrelations among serum sex hormones in postmenopausal women. *Am. J. Epidemiol.*, 133, 50–57

Fleiss J.L. (1986) *The Design and Analysis of Clinical Experiments*, New York, John Wiley & Sons, pp. 1–5

Garewal, H.S., Ramsey, L., Kaugars, G. & Boyle, J. (1993) Clinical experience with the micronucleus assay. *J. Cell Biochem.*, 17 (Suppl. F), 206–212

Hankinson, S.E., Manson, J.E., Spiegelman, D., Willett, W.C., Longcope, C. & Speizer, F.E. (1995) Reproducibility of plasma hormone levels in postmenopausal women over a 2–3-year period. *Cancer Epidemiol. Biomarkers Prev.*, 4, 649–54

Hill, M.J. (1991) From adenoma to carcinoma of the colorectum. *Recent Results Cancer Res.*, 122, 71–84

Lipkin, M., Blattner, W.A., Gardner, E.J., Burt R.W. Lynch, H., Deschner, E., Winawer, S. & Fraumeni, J.F. (1984) Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [^{3}H]thymidine labeling patterns in colonic epithelial cells. *Cancer Res.*, 44, 4201–4207

Lipkin, M., Uehara, K., Winawer, S., Sanchez, A., Bauer, C., Phillips, R., Lynch, H.T., Blattner, W.A. & Fraumeni, J.F., Jr (1985) Seventh-Day Adventist vegetarians have a quiescent proliferative activity in colonic mucosa. *Cancer Lett.*, 26, 139–144

Lyles, C.M., Sandler, R.S., Keku, T.O., Kupper, L.L., Millikan, C., Murray, S.C., Bangdiwala, S.I. & Ulshen, M.H. (1994) Reproducibility and variability of the rectal mucosal proliferation index using proliferating cell nuclear antigen immunohistochemistry. *Cancer Epidemiol. Biomarkers Prev.*, 3, 597–605

Muto, T., Bussey, H.J.R. & Morson, B.C. (1975) The evolution of cancer of the colon and rectum. *Cancer*, 36, 2251–2270

Prentice, R.L. (1989) Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.*, 8, 431–440

Prentice, R., Thompson, D., Clifford, C., Gorbach, S., Goldin, B. & Byar, D. (1990) Dietary fat reduction and plasma estradiol concentration in healthy premenopausal women. *J. Natl Cancer Inst.*, 82, 129–134

Reichman, M.E., Judd, J.T., Longcope, C., Schatzkin, A., Nair, P.P., Campbell, W.S., Clevidence, B.A. & Taylor, P.R. (1993) Effects of moderate alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. *J. Natl Cancer Inst.*, 85, 722–727

Schatzkin, A., Freedman, L.S., Schiffman, M.H. & Dawsey, S.M. (1990) The validation of intermediate endpoints in cancer research. *J. Natl Cancer Inst.*, 82, 1746–1752

Schatzkin, A., Lanza, E., Freedman, L.S., Tangrea, J., Cooper, M.R., Marshall, J.R., Murphy, P.A., Selby, J.V., Shike, M., Schade, R.R., Burt, R.W., Kikendall, W. & Cahill, J. (1996) The Polyp Prevention Trial: I. rationale, design, recruitment, and baseline participant characteristics. *Cancer Epidemiol. Biomarkers Prev.*, 5, 375–384

Schiffman, M.H. (1992) Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. *J. Natl Cancer Inst.*, 84, 394–398

Schiffman, M.H., Bauer, H.M., Hoover, R.N., Glass, A.G., Cadell, D.M., Rush, B.B., Scott, D.R., Sherman, M.E., Kurman, R.J., Wacholder, S., Stanton, C.K. & Manos, M.M. (1993) Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J. Natl Cancer Inst.*, 85, 958–964

Schiffman, M.H. & Schatzkin, A. (1994) Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res.*, 54, 1944s–1947s

Toniolo, P., Koenig, K.L., Pasternack, B.S., Banerjee, S., Rossenberg, C., Shore, R.E., Strax, P. & Levitz, M. (1994) Reliability of measurements of total, protein-bound, and unbound estradiol in serum. *Cancer Epidemiol. Biomarkers Prev.*, 3, 47–50

Toniolo, P.G., Levitz, M., Zeleniuch-Jacquotte, A., Banerjee, S., Koenig, K.L., Shore, R.E., Strax, P. & Pasternack, B.S. (1995) A prospective study of endogenous estrogens and breast cancer in postmenopausal women. *J. Natl Cancer Inst.*, 87, 190–197

Corresponding author

A. Schatzkin

Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD 20892, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Biomarker end-points in cancer chemoprevention trials

C.W. Boone and G.J. Kelloff

Over the last decade, the Chemoprevention Branch, Division of Cancer Prevention and Control, National Cancer Institute, USA, has been developing drugs that will slow or stop the progression to invasive cancer of precancerous (pre-invasive) lesions generally termed 'intraepithelial dysplasia' or 'dysplasia'. Over 40 short-term clinical trials are in progress, testing the following classes of agents on precancerous lesions in the different major organ systems: antimutagens (*N*-acetylcysteine, oltipraz), antiproliferatives (difluoromethylornithine, dehydroepiandrosterone, selenomethionine), antioxidants (vitamin E, curcumin), anti-inflammatories (aspirin, piroxicam, ibuprofen, sulindac sulfone) and hormonally active agents (tamoxifen in breast ductal carcinoma *in situ* and finasteride in prostatic intraepithelial neoplasia). Because of the strong practical need to keep so many clinical trials as short-term as possible, certain tissue changes known to be associated with high cancer risk were selected for use as biomarker end-points in the trials, such changes being quantitatively assayed by computer-assisted image analysis. These 'surrogate end-point biomarkers' (SEBs) are based on the individual cellular morphological and functional changes universally used by histopathologists to diagnose the lesion of intraepithelial neoplasia (Riddell, 1984; Boone *et al.*, 1992; Wright *et al.*, 1994). High grades of this lesion precede invasive cancer in the great majority of cases, and therefore SEBs based on them are linked to high cancer risk. Table 1 summarizes some of the short-term clinical trials now being monitored by the Chemoprevention Branch. The SEBs abbreviated 'PPNN' in the figure are: proliferative index (P); ploidy (DNA histogram) (P); nuclear morphometry and chromatin texture (N); and nucleolar size and frequency (N). Computer-assisted image analysis is used to assay these features quantitatively, which gives the SEBs increased objectivity, reproducibility and sensitivity. Further details concerned with cancer chemoprevention trials using SEBs, and their relation to the field of cancer epidemiology, are given below.

Cancer chemoprevention in relation to cancer epidemiology

Rothman (1986) describes clinical trials as a type of experimental epidemiology, stating that clinical trials, where the word 'trial' is used as a synonym for 'experiment', are epidemiological studies of different treatments for patients who already have the disease, and that 'the exposures in a clinical trial ... are preventives of the sequelae of the initial disease'. For this presentation, the specialized area of experimental epidemiology concerned with clinical trials will be called 'clinical epidemiology' and contrasted with what will be called traditional 'environmental epidemiology'. This contrast is described in Table 2.

Whereas environmental cancer epidemiology focuses on the cause and pathogenesis of cancer, clinical cancer epidemiology focuses on the diagnosis and therapy of cancer, and particularly on the conduct of randomized prospective intervention trials of therapeutic agents. Both areas of cancer epidemiology focus on cancer prevention, but from different perspectives. The environmental cancer epidemiologist favours 'primary prevention' and seeks to proscribe elements in the environment that are known to be associated with increased cancer risk, e.g. tobacco and alcohol use, exposure to fossil fuel combustion products, and fatty diet. The clinical cancer epidemiologist favours 'secondary prevention' of already established pre-invasive neoplastic disease and seeks to prescribe drugs that will prevent progression to the stage of invasiveness. Although chemoprevention is characterized here as a form of secondary prevention, in that it prevents the further sequelae (i.e. invasive cancer) of the 'primary disease' (i.e. neoplasia), it could be

Table 1. Summary of some of the short-term phase II cancer chemoprevention trials using SEBs being monitored by the Chemoprevention Branch

Organ	Chemoprevention	Gross lesion (Microlesion = dysplasia)	SEBs	Other SEBs
Skin	4HPR	Actinic keratoses	PPNN[a]	PCNA, RAR
Oral	4HPR	Leukoplakia	PPNN	PCNA, RAR, EGFR, TGF-β, involucrin
Breast	4HPR	Mammogram: ductal carcinoma *in situ*	PPNN	PCNA, *p53*, *erbB-2*
Lung	4HPR	Chronic smoker	PPNN	PCNA, EGFR, *p53*, FISH aneusomy
Bladder	DFMO	Focus of redness	PPNN	ODC, EGFR, Le^x
Colon	Ca^{2+}	Adenomatous polyp	PPNN	TThy, BrdU, extLe^x, keratins, integrins
Prostate	Proscar	PIN[b]	PPNN	
Cervix	4HPR	Mosaicism, punctation	PPNN	DNA aneuploidy, PCNA, K-*ras*

[a]PPNN: proliferative index; ploidy; nuclear morphometry; nucleolar morphometry.
[b]PIN: prostatic intraepithelial neoplasia.

argued that chemoprevention has elements of primary prevention because it involves arresting the neoplastic process before invasive cancer develops. Regardless of this point, it is clear that cancer chemoprevention is an activity that belongs to the area of clinical cancer epidemiology.

The relationship between cancer chemoprevention and molecular cancer epidemiology

The rapidly expanding field of molecular cancer epidemiology was developed using the terminology of environmental epidemiology (Hulka *et al.*, 1990; Perera *et al.*, 1982). Figure 1 illustrates how the biomarkers identified along the carcinogen exposure–neoplastic disease sequence were defined: (1) the external (carcinogen) dose biomarker; (2) the internal (carcinogen) dose biomarker; (3) the biologically effective dose (BED) biomarker (e.g. DNA and haemoglobin adducts to carcinogen); and (4) the biological response (BR) biomarker (e.g. stable mutations produced by a carcinogen). The (BR) biomarker may be considered as an early molecular lesion at the beginning of the sequence of neoplastic disease progression [the precedent for calling an abnormal molecular change a 'lesion' has been established by Ames (1995), who makes a useful distinction between DNA lesions, such as strand breaks and carcinogen adducts (BED markers), and DNA mutations (BR markers), which require cell division for their production from DNA lesions]. The fifth biomarker in this sequence has been called the subclinical disease marker (subclinical in the sense that the subject with the marker has no signs or symptoms), and is defined by Hulka *et al.* (1990) as 'a measurable indicator of a stage of disease or a manifestation of the disease itself'.

Clearly, as shown in Table 2, the subclinical diagnostic disease markers related to intraepithelial neoplasia (dysplasia) which are used as surrogate endpoint biomarkers (SEBs) in clinical trials of chemopreventive agents are the same subclinical

disease markers as defined by Hulka *et al.* (1990). They are described more fully below.

Intraepithelial (pre-invasive) neoplastic disease: the target of cancer chemoprevention

Multifocal intraepithelial (pre-invasive) neoplastic disease, a lesion of which is illustrated in Fig. 2, is the target lesion for the programme of the Chemoprevention Branch at the US National Cancer Institute. For most organ epithelia, in particular breast (Page *et al.*, 1985), prostate (Bostwick, 1992), colon (The Multicentric Study of Colorectal Adenomas Workgroup, 1995), cervix (Chanen, 1990) and lung (Carter, 1985), multifocal lesions of the intraepithelial neoplasia arise in the epithelial compartment as monoclonal expansions at multiple sites and progress at different rates for many years. Some intraepithelial neoplastic lesions remain stable or, in earlier stages, even regress. In a small proportion of lesions, however, the slowly growing intraepithelial neoplastic cell populations continue to progress to the point of invasion across the basement membrane, at which time the lesion is, by definition, diagnosed as cancer. Over the many years it takes intraepithelial neoplasia to progress, the goal of cancer chemoprevention is to find drugs that will drastically slow, stop or bring about regression of the intraepithelial neoplastic lesions while they are still confined to the epithelial compartment, thereby preventing progression of any one of them to invasive cancer.

Validation that reduction in the extent of Intraepithelial neoplasia (dysplasia) is a SEB that predicts a reduction in cancer incidence

In epidemiological terms, intraepithelial neoplasia is a necessary component on the final common pathway to cancer. The concept is straightforward—if one were to give a chemopreventive drug that eradicated all of the multifocal lesions of intraepithelial neoplasia (dysplasia), one could expect to have eliminated the possibility of the subsequent development of cancer (Fig. 2). Thus, it appears reasonable that reduction of the multiple lesions of intraepithelial neoplasia, as determined by visual inspection and multiple biopsies, may not be difficult to validate as a marker that predicts a reduction in cancer incidence. Although multifocal intraepithelial neoplasia (dysplasia) is a necessary condition permitting further progression to cancer, it may not be a sufficient one, since the majority of dysplastic epithelial lesions do not progress to invasion and may even regress.

SEBs based on the diagnostic morphological features of intraepithelial neoplasia (dysplasia) measured by computer-assisted quantitative image analysis (CQIA)

Proliferative status. The continuous and abnormal increase in the proliferative rate of intraepithelial neoplastic lesions is part of the process of clonal evolution, which is described as the continuous production within the neoplastic population of genetic variants that are able to escape growth

Table 2. Contrasting features of traditional environmental cancer epidemiology and clinical cancer epidemiology

Environmental cancer epidemiology	Clinical cancer epidemiology
Epidemiologists: toxicologists	Oncologists, pathologists
Focus: cause and pathogenesis of cancer	Focus: diagnosis and therapy of cancer
Observational studies	Prospective intervention trials
Cancer prevention by proscription (primary prevention)	Cancer prevention by prescription (secondary prevention)
Biomarkers: Biologically effective dose (BED) e.g. adducts Biological response (BR), e.g. mutations measured individually by computer-assisted quantitative image analysis	Biomarkers: subclinical disease markers Pap smear showing nuclear atypia Biopsy showing diagnostic disease markers for dysplasia measured individually by computer-assisted quantitative image analysis

inhibitory controls and undergo clonal expansion at the expense of the more slowly growing background cells (Nowell, 1986; Boone *et al.*, 1992 (review), 1993). As selected clones in the population continue to arise and expand by virtue of their faster growth rate, the overall mean growth rate of the neoplastic population increases.

The proportion of tissue cells in the proliferative cycle versus those in G_0 (not proliferating) can be accurately measured using CQIA. The computer measures the percentage of the total nuclear area in a section that binds to a chromogen-labelled antibody probe, such as antibodies to PCNA, Ki-67, (using MIB-1) or BrdU (technique described in Esteban *et al.*, 1993). The proliferating fraction may also be measured by tritiated-thymidine uptake/autoradiography or by mitotic counts. The mitotic index (percentage of cells in mitosis) has proven to be a reliable prognostic factor in breast cancer, accurately predicting the recurrence-free survival and overall survival either by itself (van Diest *et al.*, 1992) or as part of a commonly used grading system (Bloom, 1950). In the colorectum, an upward shift of the zone of maximal proliferation towards the neck of the gland ('stage II anomaly') has been shown to be an important predictor of increased cancer risk (Deschner & Haskens, 1982).

Ploidy status. The DNA content per cell nucleus measured by CQIA is displayed as a histogram, from which the type and degree of variation from diploidy may be quantitatively estimated. Aneuploidy has been shown to occur during intraepithelial neoplasia in bladder (Tribukait, 1987), prostate (Montironi *et al.*, 1990; Tribukait, 1991; Amin *et al.*, 1993), breast (Carpenter *et al.*, 1987; Erhardt & Auer, 1987; Crissman *et al.*, 1990;

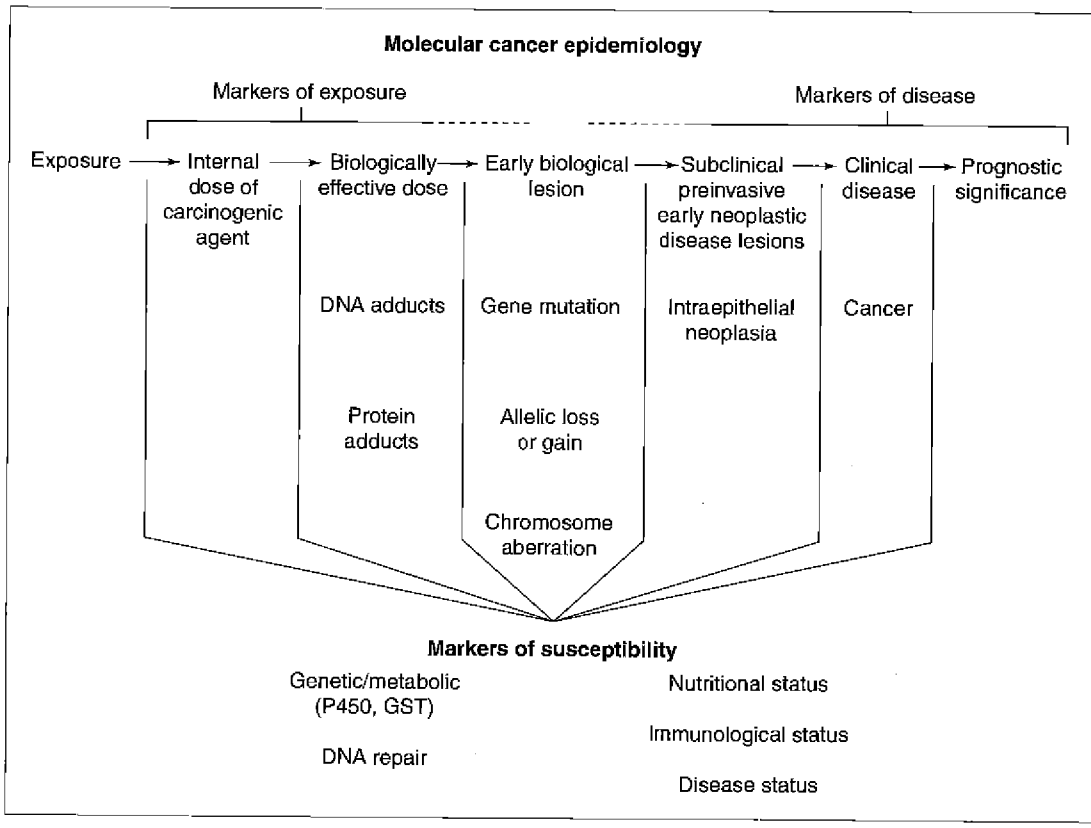

Figure 1. The position of intraepithelial neoplasia (dysplasia), shown as a subclinical disease biomarker, in the carcinogen exposure–neoplastic disease sequence.

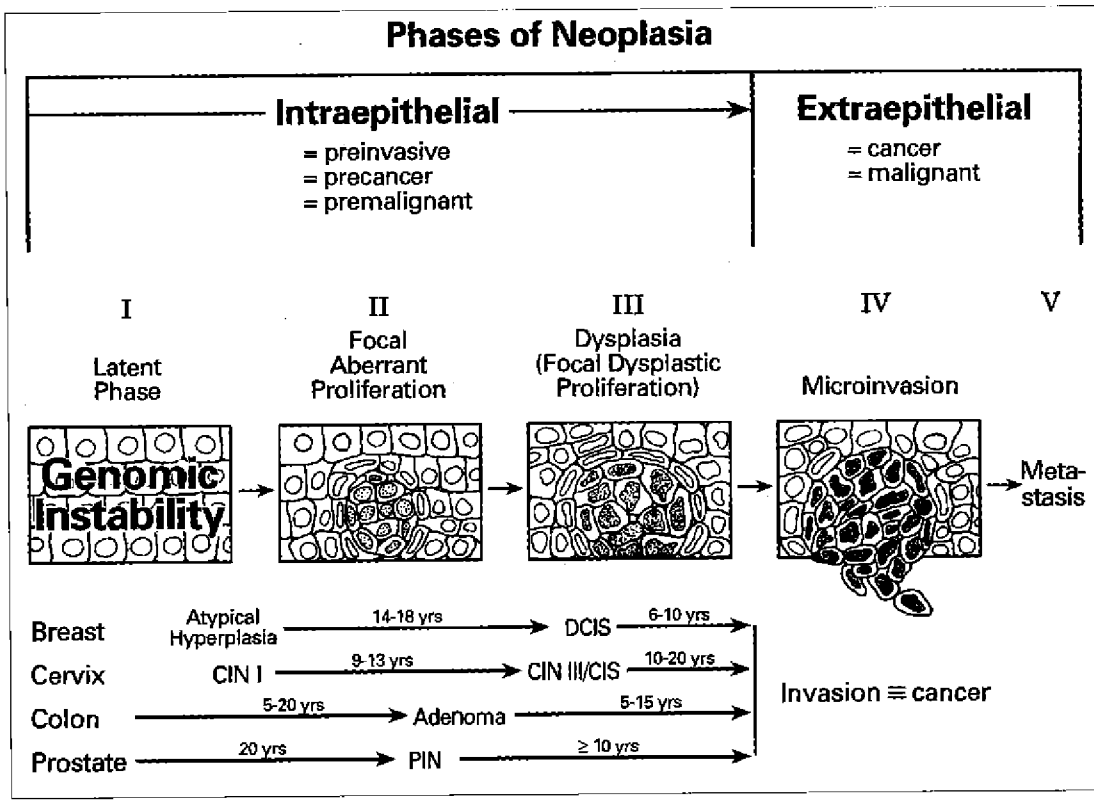

Figure 2. Diagram illustrating the concept that if multifocal intraepithelial neoplasia is stopped by the administration of a chemopreventive drug, further progression to invasive cancer is prevented.

Visscher *et al.*, 1993), cervix, skin, oral leukoplakia, larynx, lung, oesophagus, stomach and colorectum (reviewed in Boone *et al.*, 1992). In one study of breast ductal carcinoma *in situ*, the cribriform histological pattern exhibited 38% aneuploidy, whereas the comedo pattern exhibited 82% aneuploidy (Crissman *et al.*, 1990). In another study, atypical hyperplasia of the breast also exhibited aneuploidy in 4 of 13 cases (Carpenter *et al.*, 1987). With regard to invasive neoplasia, ploidy status has proven to be a reliable prognostic factor of disease-free survival and mortality in both breast (van Diest & Baak, 1992) and prostate (Lieber, 1992) cancer. Concerning pre-invasive neoplasia (dysplasia), DNA aneuploidy determined by CQIA has proven to be an effective SEB that correlates with the severity of intraepithelial neoplasia of breast (Beerman *et al.*, 1990) and prostate (Montironi *et al.*, 1990; Lieber, 1992; Petein *et al.*, 1992; Amin *et al.*, 1993).

Nuclear morphometry (nuclear size, shape and pleomorphism). It is quite remarkable that, in a number of studies, alteration of nuclear shape alone proved to be a better predictor of mortality in stage A2 prostatic cancer than were the Gleason, Mostofi or Johns Hopkins grading systems, or ploidy status (Epstein *et al.*, 1984; Partin *et al.*, 1989; Mohler *et al.*, 1992;). Multivariate analysis of up to 16 nuclear shape descriptors, measured by CQIA, including nuclear roundness factor, variance of roundness factor and nuclear ellipticity, have accurately predicted recurrence of cancer after surgery in 11 of 26 patients with renal cell carcinoma (Pound *et al.*, 1993), 7/14 patients with transitional cell carcinoma of the bladder (Borland *et al.*, 1993), and 17/27 patients with Wilms' tumour of the kidney (Gearhart *et al.*, 1992). Nuclear pleomorphism is an important concomitant of neoplasia that may be measured as the simple mean of the variances of size and shape.

Nuclear chromatin texture. 'Chromatin clumping' is a semiquantitative nuclear feature long recognized by pathologists as a correlate of the extent of neoplastic progression. Mildly dysplastic cells, for instance, tend to exhibit less chromatin clumping than severely dysplastic cells. With CQIA, the degree of chromatin clumping may be measured precisely. The multitude of possible optical density patterns transmitted through many hundreds of pixels over a given cell nucleus has been systematized into what is known as the Markovian texture feature matrix (Pressman, 1976). Selection may be made among dozens of different Markovian texture features for those which best contribute to a multivariate classification function that will predict the likelihood of intraepithelial neoplasia progressing to invasive cancer (Doudkine *et al.*, 1995). Nuclear chromatin texture analysis by CQIA therefore has excellent potential for providing many features which are useful as SEBs.

Nucleolar morphometry (number, size, shape, position and pleomorphism). In a study of nine nucleolar morphometrical features in breast cancer, simple nucleolar frequency among nuclei (total number of nucleoli per 100 nuclei) was a significant predictor of recurrence-free survival (van Diest *et al.*, 1990). Changes in nucleolar morphometry have been reported to be a correlate of the extent of neoplastic progression in prostatic intraepithelial neoplasia (Helpap, 1988; Montironi *et al.*, 1991). Argyrophilic nucleolar organizer region-associated proteins found within nucleoli form the basis for the AgNOR stain. Using this stain, the mean number of nucleolar organizer regions (NORs) clearly distinguish between tubular adenomas, villous adenomas with moderate nuclear atypia, villous adenomas with severe nuclear atypia and colorectal adenocarcinoma (Yang *et al.*, 1990). The AgNOR stain, especially if quantitated by CQIA, offers good potential as a SEB.

Continuous grading scale measured in standard deviation units. The individual morphological and proliferative features of intraepithelial neoplasia discussed above characteristically progress as a group towards greater aberrancy. The quantitative extent of aberrant variation of each feature in the group, measured by computer-assisted image analysis, may be graded in terms of the number of standard deviation units from the mean, and a critical value diagnostic of neoplasia may be defined for each feature (Dr James Bacus, personal communication). A composite biomarker made up of a number of diagnostic features may then be constructed.

Development of molecular SEBs

The Chemoprevention Branch is also evaluating various candidate molecular SEBs as predictors of high cancer risk and for their efficacy in short-term clinical trials. Molecular SEBs being studied include those related to genomic instability, oncogene amplification, tumour suppressor gene loss, aberrant differentiation molecules and aberrant expression of growth regulatory molecules and their receptors. There are screening programmes in the planning stage that will allow the collection of cohorts of subjects with activated oncogenes or mutated or lost tumour suppressor genes, or with genetic markers of susceptibility to cancer, which may become the subject of chemoprevention trials.

The most desirable molecular SEBs may be those that appear before the onset of intraepithelial neoplasia in normal-appearing epithelium. The evaluation of early genomic instability is an example. The development of such early 'pre-dysplastic' SEBs, which is now in progress in the programme of the Chemoprevention Branch, will permit many more months or even years of treatment with chemopreventive agents over a period when the molecular lesions may be more susceptible to suppressive drugs. Molecular SEBs that are superimposed on neoplastic tissue changes already identifiable microscopically may offer considerable additional information in tissue biopsies, particularly as technology involving microdissection of individual cells followed by PCR and gel analysis becomes established.

References

Ames, B.N. (1995) The causes and prevention of cancer. *Proc. Natl Acad. Sci. USA*, 92, 5258–5265

Amin, M.B., Schultz, D.S., Zarbo, R.J., Kubus, J. & Shaheen, C. (1993) Computerized static DNA ploidy analysis of prostatic intraepithelial neoplasia. *Arch. Pathol. Lab. Med.*, 117, 794–798

Beerman, H., Kluin, P.M., Hermans, J., Velde, C.J.H. & Cornelisse, C.J. (1990) Prognostic significance of DNA-ploidy in a series of 690 primary breast cancer patients. *Int. J. Cancer*, 45, 3–39

Bloom, H.J.G. (1950) Prognosis in carcinoma of the breast. *Br. J. Cancer*, 4, 259–288

Boone, C.B., Kelloff, G.J. & Steele, V.E. (1992) Natural history of intraepithelial neoplasia in humans with implications for cancer chemoprevention strategy. *Cancer Res.*, 52, 1651–1659

Boone, C.W., Kelloff, G.J. & Freedman, L. (1993) Intraepithelial and postinvasive neoplasia as a stochastic continuum of clonal evolution, and its relationship to mechanisms of chemopreventive drug action. *J. Cell Biochem.*, 17G, 14–25

Borland, R.N., Partin, A.W., Epstein, J.I. & Brendler, C.B. (1993) The use of nuclear morphometry in predicting recurrence of transitional cell carcinoma. *J. Urol.*, 149, 272–275

Bostwick, D.G. (1992) Prostatic intraepithelial neoplasia (PIN): current concepts. *J. Cell Biochem.*, 16H, 10–19

Carpenter, R., Gibbs, N., Matthews, J. & Cooke, T. (1987) Importance of cellular DNA content in pre-malignant breast disease and pre-invasive carcinoma of the female breast. *Br. J. Surg.*, 74, 905–906

Carter, D. (1985) Squamous cell carcinoma of the lung: an update. *Semin. Diag. Pathol.*, 4, 226–234

Chanen, W. (1990) The CIN saga, the biological and clinical significance of cervical intraepithelial neoplasia. *Aust. NZ. J. Obstet. Gynaecol.*, 30, 18–23

Crissman, J.D., Visscher, M.D. & Kubus, J. (1990) Image cytophotometric DNA analysis of atypical hyperplasia and intraductal carcinomas of the breast. *Arch. Pathol. Lab. Med.*, 114, 1249–1253

Deschner, E.E. & Maskens, A.P. (1982) Significance of the labeling index and labeling distribution as kinetic parameters in colorectal mucosa of cancer patients and DMH treated animals. *Cancer*, 50, 1136–1141

Doudkine, A., MacAulay, C., Poulin, N. & Palcic, B. (1995) Nuclear texture measurements in image cytometry. *Pathologica*, 87, 180–193

Epstein, J.I., Berry, S.J. & Eggleston, J.C. (1984) Nuclear roundness factor. A predictor of progression in untreated Stage A_2 prostate cancer. *Cancer*, 54, 1666–1671

Erhardt, K. & Auer, G.U. (1987) Mammary carcinoma. Comparison of nuclear DNA content from *in situ* and infiltrative components. *Anal. Quant. Cytol. Histol.*, 9, 263–267

Esteban, J.M., Kandalaft, P.L., Mehta, P., Odom-Maryon, T.L., Bacus, S. & Battifora, H. (1993) Improvement of the quantitation of estrogen and progesterone receptors in paraffin-embedded tumors by image analysis. *Am. J. Clin. Pathol.*, 99, 32–38

Gearhart, J.P., Partin, A.W., Leventhal, B. & Beckwith, J.I. (1992) The use of nuclear morphometry to predict response to therapy in "Wilms" tumor. *Cancer*, 69, 804–808

Helpap, B. (1988) Observations on the number, size, and localization of nucleoli in hyperplastic and neoplastic prostatic disease. *Histopathology*, 13, 203–211

Hulka, B.S., Wilcosky, T.C. & Griffith, J.D. (1990) *Biologic Markers in Epidemiology*, New York, Oxford University Press

Lieber, M.M. (1992) DNA ploidy: early malignant lesions. *J. Cell Biochem.*, 16H, 44–46

Mohler, J.L., Partin, A.W., Epstein, J.I., Becker, R.L., Mikel, U.V., Sesterhenn, I.A., Mostofi, F.K., Gleason, D.F., Sharief, Y. & Coffey, D.S. (1992) Prediction of prognosis in untreated stage A_2 prostatic carcinoma. *Cancer*, 69, 511–519

Montironi, R., Scarpelli, M., Sisti, S., Braccischi, A., Gusella, P., Pisani, E., Alberti, R. & Mariuzzi, G.M. (1990) Quantitative analysis of prostatic intraepithelial neoplasia on tissue sections. *Anal. Quant. Cytol. Histol.*, 12, 366–372

Montironi, R., Braccischi, A., Matera, G., Scarpelli, M. & Pisani, E. (1991) Quantitation of the prostatic intraepithelial neoplasia. Analysis of the nucleolar size, number, and location. *Pathol. Res. Pract.*, 187, 307–314

Nowell, P.C. (1986) Mechanisms of tumor progression. *Cancer Res.*, 46, 2203–2207

Page, D.L., Dupont, W.D., Rogers, L.W. & Rados, M.S. (1985) Atypical hyperplastic lesions of the female breast: a long-term follow-up study. *Cancer*, 55, 2698–2708

Partin, A.W., Walsh, A.C., Pitcock, R.V., Mohler, J.L., Epstein, J.I. & Coffey, D.S. (1989) Comparison of nuclear morphometry and Gleason grade as predictor of prognosis in Stage A_2 prostate cancer: a critical analysis. *J. Urol.*, 142, 1254–1258

Perera, F.P. & Weinstein, I.B. (1982) Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J. Chronic Dis.*, 35, 581–600

Petein, M., Michel, P., van Velthoven, R., Pasteels, J.L., Brawer, M.K., Davis, J.R., Nagle, R.B. & Kiss, R. (1991) Morphonuclear relationship between prostatic intraepithelial neoplasia and cancer as assessed by digital cell image analysis. *Am. J. Clin. Pathol.*, 96, 628–634

Pound, C.R., Partin, A.W., Epstein, J.I., Simons, J.W. & Marshall, F.F. (1993) Nuclear morphometry accurately predicts recurrence in clinically localized renal cell carcinoma. *Urology*, 42, 243–248

Pressman, N.J. (1976) Markovian analysis of cervical cell images. *J. Histochem. Cytochem.*, 24, 138–144

Riddell, R.H. (1984) Dysplasia and cancer in ulcerative colitis: a soluble problem? *Scand. J. Gastroenterol.*, 104, 137–149

Rothman, K.J. (1986) *Modern epidemiology*. Boston, Little, Brown. pp. 52–53

The Multicentric Study of Colorectal Adenomas (SMAC) Workgroup (1995) A multicenter study of colorectal adenomas: rationale, objectives, methods, and characteristics of the study cohort. *Tumori*, 81, 157–163

Tribukait, B. (1987) Flow cytometry in assessing the clinical aggressiveness of genito-urinary neoplasms. *World J. Urol.*, 5, 108–122

Tribukait, B. (1991) DNA flow cytometry in carcinoma of the prostate for diagnosis, prognosis, and study of tumor biology. *Acta Oncol.*, 30, 187–192

van Diest, P.J., Mouriquand, J., Schipper, N.W. & Baak, J.P.A. (1990) Prognostic value of nucleolar morphometric variables in cytological breast cancer specimens. *J. Clin. Pathol.*, 43, 157–159

van Diest, P.J. & Baak, J.P.A. (1992) *Quantitative Cyto- and Histoprognosis in Breast Cancer*, New York, Elsevier, pp. 55–62

van Diest, P.J., Baak, J.P.A., Matze-Cok, P. & Wisse-Brekelmans, E.C.M. (1992) Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the multicenter morphometric mammary carcinoma project. *Hum. Pathol.*, 23, 603–607

Visscher, D.W., Micale, M.A. & Crissman, J.D. (1993) Pathological and biological relevance of cytophotometric DNA content to breast carcinoma genetic progression. *J. Cell Biochem.*, 17G, 114–122

Wright, T.C., Kurman, R.J. & Ferenczy, A. (1994) Precancerous lesions of the cervix. In: Kurman, R.J., ed., *Blaustein's Pathology of the Female Genital Tract*, New York, Springer-Verlag

Yang, P., Huang, G.S. & Zhu, X.S. (1990) Role of nucleolar organizer regions in differentiating malignant from benign tumours of the colon. *J. Clin. Pathol.*, 43, 235–238

Corresponding author
C.W. Boone
Chemoprevention Branch, Division of Cancer Prevention and Control, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

The use of biological markers as predictive early-outcome measures in epidemiological research

A.J. McMichael and A.J. Hall

One of the possible uses of biomarkers in epidemiological research is as early-outcome measures to predict the occurrence of clinical disease and to elucidate the biological mechanism of pathogenesis. This use is conceptually less straightforward than the well established use of biomarkers to improve or extend exposure assessment or to study interindividual variations in disease susceptibility. In principle, this form of use could accelerate or otherwise facilitate etiological research. However, in practice, the recent review literature suggests that this mode of biomarker use, especially in cancer epidemiology, is the least clear-cut and the least well developed. The recurrent problem is identifying biomarkers that: (1) are on the causal pathway, (2) have a high probability of progression to clinical disease, and (3) account for all or most of the cases of the specified clinical outcome. Such biomarkers would be most useful if they conferred a long lead-time relative to clinical disease occurrence.

The use of biological markers, molecular or otherwise, in epidemiological research should be as a means and not as an end (Hulka, 1991; McMichael, 1994). Therefore, the critical question to ask in any particular study is: 'Has use of the biomarker in this study yielded new information, or better information, for understanding, and perhaps quantifying, the causation of the disease of interest?'

The four major contexts in which biomarkers may be used in epidemiological research are: (1) to improve the assessment of exposure; (2) to identify and take account of subgroups of persons of differing susceptibility to the effects of the exposure; (3) as measures of early outcome with known (or presumed) predictive significance; and (4) as a basis for differentiating disease subtypes with potentially different etiologies.

The first of these uses, improving the assessment of exposure, can be enhanced by use of a wide range of biological markers, from blood cholesterol and urinary cotinine, to levels of DNA adducts in surrogate and target tissue cells. Some such measures will identify component exposures within a complex mixture, others will confirm that an internal exposure has indeed occurred, while others yield an integrated measure of 'biologically effective dose'. This last integrates either multiple external sources of exposure or multiple internal metabolic steps, or both.

The second mode of biomarker use—the identification of subgroups of study subjects with different levels of susceptibility to some particular exposure—often depends on measuring some biochemical or molecular genetic characteristic. For example, one recent study indicated that the reduction in coronary heart disease mortality risk associated with increasing levels of consumption of alcohol, within the low–moderate range, depends on whether the individual is Lewis blood group positive or negative (Hein *et al.*, 1993). Similarly, the risk of colon cancer associated with increasing levels of meat consumption appears to depend on whether an individual is a fast or a slow acetylator (Roberts-Thomson *et al.*, 1996)—a polymorphic metabolic characteristic that is measurable directly via genotype analysis of *NAT2* alleles (Minchin *et al.*, 1993; Roberts-Thomson *et al.*, 1996).

These two uses of biomarkers can sometimes be combined to elucidate, or confirm, epidemiological relationships. For example, cigarette smokers are known to be at increased risk of cancer of the urinary bladder. The risk of this cancer is also higher in

slow acetylators than in fast acetylators (Lower *et al.*, 1979; Hanssen *et al.*, 1985; Cartwright *et al.*, 1982). Since arylamines, present in cigarette smoke, are a known occupational cause of bladder cancer (especially β-naphthylamine) and are metabolized by *NAT2*, the acetylation pathway, it has been postulated that the carcinogenic effect of certain arylamines might account for the epidemiological findings. Studies of the levels of DNA adducts in white blood cells in different categories of persons have shed further light on those relationships: smokers have higher levels of 4-amino-biphenyl-haemoglobin adducts than non-smokers (Bryant *et al.*, 1987), and in smokers (especially 'passive smokers') the level is higher in slow than in fast acetylators (Vineis *et al.*, 1994).

The third possible use of biomarkers—the use of early-outcome measures to predict the occurrence of clinical disease—is the subject of this chapter. It is conceptually less straightforward than the previous two uses; its principal purposes appear to be to accelerate (or otherwise facilitate) etiological epidemiological research and to elaborate the intervening steps along the causal pathway. The recent review literature suggests that this third mode of biomarker use, especially in cancer epidemiology, is the least clear-cut and the least well developed (Perera & Santella, 1993; McMichael, 1994). Typically, there is a preference for identifying and using biomarkers that confer a long lead-time relative to the appearance of clinical disease. Such biomarkers may be cellular, biochemical or molecular.

The fourth above-mentioned use of biomarkers is outside the scope of this paper. By differentiating disease subtypes and thus increasing the specificity of the data analysis, this particular use should enhance the informativeness of a study (e.g. Taylor *et al.*, 1992; Vahakangas *et al.*, 1992).

Uses of 'predictive' biomarkers

There are four main uses of biomarkers as predictors (or 'surrogate end-points'): (1) screening for pre-clinical disease; (2) enhancement of conventional epidemiological studies of disease etiology; (3) monitoring for variation or change in population health risk; and (4) confirmation, via controlled intervention studies, of biomarkers that could subsequently be applied in the evaluation and monitoring of primary prevention measures.

In each of the first three of these uses, the predictive ability of the biomarker is inferred from knowledge of the natural history of the disease, and is then applied in ways that save time in epidemiological research or that save time (and lives) in public health prevention. In the fourth use, one examines whether the inferred predictivity of the biomarker is also evident under the strict conditions of the controlled intervention—i.e. is there a strong within-individual correlation between changes in the biomarker, as induced by the intervention, and occurrence of the clinical end-point (e.g. the Concorde Trial; see section on the 'confirmation of the predictivity of biomarkers' below)?

Routine screening for (immediately) pre-clinical conditions

There is a well established use of precursor lesions at the macroscopic level in clinical medicine for the early detection of disease, e.g. precancerous polyps of the large bowel and hypertension-associated arteriovenous 'nipping' within the retina. Similarly, the early detection of women likely to develop cancer of the cervix is achieved by taking cell samples from the cervical mucosa and seeking microscopic evidence of dysplasia. In general, this type of relatively late-stage precursor lesion is useful for population screening programmes directed at the secondary prevention of clinical disease.

A more novel cancer screening idea has recently arisen from the observation that persons with adenomatous polyps (which carry a high probability of progression to carcinoma) may be identified by the presence, in shed cells in faecal material, of the K-*ras* (oncogene) mutation (Sidransky *et al.* 1992). This mutation is one of the well known four mutations and is thought to be critical in the development of colon carcinoma (Vogelstein *et al.*, 1988; Fearon & Vogelstein, 1990). The four mutations are of the *APC* tumour suppressor gene (chromosome 5), the K-*ras* oncogene (chromosome 12), the *DCC* tumour suppressor gene (chromosome 18), and the *p53* tumour suppressor gene (chromosome 17).

Note, however, that a prerequisite for any successful screening procedure is prior knowledge that the lesion is clearly predictive of the subsequent occurrence of clinical disease. There would be little point in detecting cervical cellular dysplasia if it were only weakly predictive of subsequent cervical

carcinoma. Such knowledge of the capacity to predict prospectively can only come from studies of the natural history of the disease process. (Note that this is a different notion from that other central parameter of a screening test, 'positive predictive value' (PPV). The PPV is the probability that the presence of this lesion, as measured within a specified population of screened persons, signifies the true presence of the early, pre-clinical, disease process.)

The clinically oriented use of precursor lesions for screening, within either a clinical case-finding context or a systematic population screening context, is well established. This paper, however, addresses the use of biochemical and molecular biomarkers for two other purposes: the direct enhancement of epidemiological research into disease etiology and the substantially earlier detection of altered health risks in monitored populations. In each case, there would be a general preference for biomarkers as indices of rather earlier stages in disease pathogenesis than is the case with clinically oriented screening.

Direct enhancement of epidemiological research into disease etiology

The use of biomarkers as surrogate end-points in epidemiological research could, in principle, enhance the research in two main ways:

1. Such biomarkers could provide earlier answers in cohort studies by avoiding the need to await clinical disease end-points. Biomarker-based case–control studies could also yield earlier answers (in relation to the advent of some newly introduced exposure agent). This would require identification of a random sample of all persons ('cases') with the 'predictive' lesion. For example, if a new subtype of HPV were suspected of causing cervical cancer, women with dysplastic cells (detected via screening) could be compared with controls for the presence of viral subtype DNA.

2. Such biomarkers could elucidate, or confirm, the underlying biological process of disease pathogenesis. This, in turn, may assist in the interpretation of epidemiological findings and in the drawing of causal inference.

How well do these two main potential uses of predictive biomarkers work in practice?

Gaining earlier answers in prospective epidemiological studies. Biomarkers could, in principle, help to gain both earlier and better answers to prospective studies. Better answers may result by exclusion of ostensibly healthy persons who, via biomarker screening, are assessed as already having the pre-clinical form of the disease. For example, from many studies over the past several decades, it has been concluded that a low blood cholesterol concentration is not a cause of increased risk of cancer (with the possible exception of colon cancer), but is a metabolic manifestation of incipient cancer (McMichael *et al.*, 1984). It would therefore be possible to screen out from a newly recruited cohort all persons with manifestly low blood cholesterol—or at least identify them for subsequent stratified analysis of cancer incidence.

Earlier answers may be obtained from epidemiological studies if the use of a biomarker increases the statistical power of a study (McMichael 1994). For example, if instead of counting numbers of persons with cancer it were equally valid to count the number of cells from standard-sized tissue samples (e.g. 500 white blood cells per person) displaying some specific precancerous predictive mutation, then statistical power would be greatly increased (Hattis & Silver, 1993).

With respect to the task of gaining earlier answers from cohort studies, the *sine qua non* is prior knowledge that the biochemical or molecular marker is actually strongly predictive of the clinical disease, and that clinical disease never or rarely occurs without this antecedent marker, i.e. the 'attributable fraction' approaches unity. Note that this does not mean that the biomarker is necessarily on the causal pathway; the biomarker may simply be an outcome that is very highly correlated with the occurrence of the clinical disease. It is only necessary, in a logical sense, that the biomarker can be used as a surrogate, antecedent, event for the actual clinical disease event. This is nevertheless a demanding criterion: if the biomarker does not represent a stage in the causal chain, there is the possibility that, while some particular exposure agent to be studied does induce the disease, it does not induce the biomarker epiphenomenon (or vice versa).

Even assuming that the chosen predictor does indeed lie on the causal pathway, there may be other late-stage risk factors that account for the

progression from early 'lesion' to clinical event. Such factors would not be identified in studies that use the biomarker as the surrogate measure of clinical disease outcome. For example, whereas dietary fat intake may substantially account for the genesis of the underlying atherosclerotic arterial disease, the actual formation of an occlusive blood clot may be precipitated by other factors, such as cigarette smoking, deficient intake of certain micronutrients (vitamin C, folic acid, etc.) or exposure to high concentrations of airborne respirable particulates.

There is an unavoidable dilemma here. To gain the most leverage in time, the biomarker should substantially antedate the clinical end-point. However, the earlier the surrogate end-point, the less likely it is to be strongly predictive of the clinical end-point. Two recent cohort studies have examined the predictivity of cytogenetic damage—structural chromosomal aberrations, sister chromatid exchanges and micronuclei (Hagmar *et al.*, 1994; Bonassi *et al.*, 1995). In both studies there is evidence that a high level of chromosomal aberrations in white blood cells foreshadows an increased risk of cancer. However, the risk increases are modest and there is no indication that such cytogenetic indices could be used as a substitute for cancer in etiological studies.

It may be helpful also to consider several examples from the non-cancer arena. A substantial scientific literature indicates that coronary atherosclerosis, with manifest narrowing of coronary arteries, is predictive of ischaemic heart disease; that the rate of loss of lung function (e.g. FEV_1) is quantitatively predictive of life-shortening at older ages; and that impaired glucose tolerance predicts the onset of non-insulin-dependent diabetes mellitus (NIDDM). We might therefore use these subclinical disorders as 'outcomes' in etiological studies, seeking to understand better the etiology of the clinical disease. For example, in an ongoing follow-up study of a Swedish cohort, the occurrence of impaired glucose tolerance in mid-adult life is being used as a surrogate for putative subsequent clinical NIDDM (Lithell *et al.*, 1996).

There are, so far, very few good examples of the use of surrogate end-points as substitutes for the eventual clinical event. Some cancer epidemiology examples exist within the controlled trial domain. However, their focus is on cytogenetic events, including specific mutations, rather than 'later' markers of cellular abnormality. For example, some studies have used the incidence of cytogenetic damage (e.g. micronucleus formation) to evaluate the cancer-protective effect of certain interventions, such as the effect of micronutrient supplementation on the progression of oesophageal dysplasia (Muñoz *et al.*, 1987). Are there other possibilities? Would we, in carrying out a cohort study of diet and colon cancer, settle for the measurement of particular cancer-associated mutations in faecally shed cells as the index of impending carcinoma? Could a very specific mutation—for example, the third base pair mutation of codon 249 in the *p53* gene, which, in liver tumour cells, is strongly (but probably not exclusively) associated with aflatoxin exposure (Bressac *et al.*, 1991; Hsu *et al.*, 1991)—be measured in the DNA of white blood cells as a surrogate index for the presumed occurrence of the same mutation in the relevant target tissue? The answer at the moment is that we do not yet have certain and specific molecular biomarkers that would be satisfactory alternative 'outcome' measures in epidemiological studies of cancer etiology.

Discussion of the use of early predictive indices in epidemiological research tends to focus on cancer epidemiology and the associated mutational events. The genotoxic model of carcinogenesis has a clarity about it that is unusual in the realm of disease etiopathogenesis. The model assumes a 'staged' mechanism that entails the sequential cumulation of critical mutations. While genotoxicity is only part of carcinogenesis, current knowledge and theory propose that it is an important part. Nevertheless, various pre-clinical chemical changes (e.g. carcinoembryonic antigen (CEA) and alpha-fetoprotein (AFP)) may be valuable as biomarkers, particularly since they probably reflect actual cellular functional change. Although they tend to occur at later stages in carcinogenesis than genetic and chromosomal damage, and therefore offer less lead-time in etiological research, they may have greater predictive power.

Few other disease processes have the seemingly 'staged' character of carcinogenesis. In heart disease, diabetes, asthma and cerebral dementia, there are no mutation-like 'switches' thrown—or at least none that we know of. For non-cancer diseases, it is therefore much harder to imagine what the critical non-tissue-based predictive biomarkers might be. (Examples are more evident

within clinical epidemiology, in the study of the factors that influence the outcome of disease—see section below.)

Overall, and until we have much more knowledge about the significance and predictive power of these various biomarkers, it is clear that they cannot simply be substituted for the clinical endpoint of classical cancer epidemiological research. Definitive answers, for the moment (perhaps always?), must continue to depend on studies that use the occurrence or non-occurrence of cancer as the outcome variable.

Elucidating, or confirming, the steps in the underlying biological process. A perennial discussion in epidemiology concerns the sufficiency of the 'black box' approach. Do we need to know more than that a particular exposure or circumstances is empirically associated with an increased risk of disease? For the progression of scientific ideas and the potentiation of future etiological research, knowledge of the mediating mechanism is important. However, for the practical success of many public health interventions, it is actually not necessary to know anything about the mediating biological mechanisms. It may be enough, then, to know that smoking cigarettes causes lung cancer, or that drinking water from the Broad Street pump causes cholera.

However, it is becoming increasingly clear to cancer epidemiologists that an understanding of mechanism—e.g. which components of cigarette smoke cause lung cancer?—will allow further studies of interindividual variation in susceptibility. In the case of smoking, such knowledge would also assist the investigation and risk quantification associated with passive smoking. On a wider research front, epidemiologists now perceive that further understanding of the role of diet in many cancers, of alcohol in breast cancer, EMF in leukaemia and brain cancer, and ambient air pollution in lung cancer will require elaboration of the mediating mechanisms. The empirical epidemiological evidence on its own is neither sufficiently strong nor consistent to allow causal inference. It therefore becomes attractive to study the sequence of probable, or plausible, mediating biological events.

The aflatoxin and liver cancer story provides an instructive example (McMichael, 1994). Aflatoxin may increase the risk of liver cancer either by direct genotoxic action, by inducing oxidative damage, or by affecting the immune response to hepatitis B virus. In relation to the latter possibility, for example, in one study the level of aflatoxin was found to be higher in children who were carriers of the virus than in those who were not (Allen *et al.*, 1992).

The prevailing view, in line with the laboratory experimental evidence of the potent direct chemical carcinogenicity of aflatoxin, is that it probably acts as a genotoxin. Recent evidence that aflatoxin may cause an unusually specific mutation of the important *p53* gene—the 'hot spot' mutation at codon 249 in which the third base pair undergoes G to T transversion—has therefore advanced discussion of this genetic option. Mutations of the *p53* gene, both 'hot spot' and otherwise, appear to occur with equal frequency in tumours associated with persistent hepatitis B infection and in those not associated with this infection (Oda *et al.*, 1992), which suggests that *p53* mutation is not a result of hepatitis B virus infection. However, there is a high frequency of this specific mutation of *p53* (45% of tumours) in the high-incidence areas of the world, such as Mozambique and Qidong, China (Ozturk *et al.*, 1991; Sheu *et al.*, 1992), where exposure to aflatoxin also tends to be raised. This ecological evidence has suggested that the mutation may be specific to genotoxic aflatoxin exposure—a notion that is further supported by the fact that this is not a 'hot spot' mutation in tumours other than hepatocellular carcinoma within those populations.

This evidence does not necessarily indicate that the *p53* mutation is the direct mechanism of aflatoxin carcinogenicity, since it may merely be acting as a surrogate marker of aflatoxin exposure in these populations. There may still be a coexistent non-genotoxic effect of aflatoxin, such as suppressing the immune defenses against HBV. However, the example illustrates how the study of mutational spectra can yield an insight into plausible causal mechanisms and, if the appropriate set of observations is made, can strengthen the basis for causal inference.

Another unresolved diet–cancer issue is whether antioxidants, especially β-carotene, reduce the risk of cancer, and if so, whether this accounts for the lowered risks of many cancers in association with above-average intake of fresh fruit and vegetables. There are evident difficulties in carrying out sufficiently large and long-term intervention trials, and inconsistent results have emerged recently from studies in China (Taylor *et al.*, 1994), Finland (ABC

Study Group, 1994) and the USA (Hennekens *et al.*, 1996). In order to help clarify this situation, Duthie and colleagues have carried out studies to assess the extent to which, first, dietary antioxidants as measured in blood are correlated with reduced reporter mutations in lymphocytes (Duthie *et al.*, 1995) and, second, antioxidant supplementation (for 20 weeks) reduces oxidative DNA damage in lymphocytes (Duthie *et al.*, 1996). The second study explicitly used biomarkers of DNA damage or mutation 'as indicators of carcinogenic risk'; the authors argued that 'DNA damage, resulting in base change and mutation when replication occurs, is a very early event in carcinogenesis'. Subsequently, the authors concluded (Duthie *et al.*, 1996):

> Herein, we show for the first time a highly significant moderating effect of long-term antioxidant supplementation on endogenous and exogenous oxidative DNA damage in lymphocytes, supporting the hypothesis that dietary antioxidants may protect against cancer.

The discourse concerning radiation-induced germline mutations as a possible cause of human cancer—highlighted by the study indicating an increased risk of leukaemia in the children of radiation-exposed Sellafield workers (Gardner *et al.*, 1990)—has recently been extended by a study in Belarus in the Mogilev district, north of Chernobyl (Dubrova *et al.*, 1996). The frequencies of lymphocyte DNA mutations (tandem repeat minisatellite mutations, detected via five separate DNA probes) were compared in samples of 79 Belarus children and 105 UK children, all born in 1994. The mutation rate in Belarus children was double that of UK children, and the rate in Belarus children showed a strong positive association with level of local topsoil radiation (caesium-137) contamination. Those mutations are presumed to have been inherited from parents who were exposed in the aftermath of Chernobyl.

This study of heritable radiation-associated mutations thus adds some plausibility to the earlier, necessarily inconclusive, epidemiological research at Sellafield. It also supports the general thesis that the dose rate is a critical determinant of risk of genetic damage, wherein the unit-dose risks are higher at lower dose rates. Concurrent research in subterranean voles living in the immediate vicinity of the Chernobyl reactor site showed a rate of mitochondrial genetic damage (base pair substitutions in the cytochrome *b* gene) that was several hundred times greater than in vole populations living 32 km south-east of Chernobyl (Baker *et al.*, 1996). It is likely that this genetic damage was also sustained over several years of low-dose-rate exposure.

Biomarkers may also be used to clarify the role of exposure agents at different stages of the cancer process. In particular, there has been long-standing interest in the notion of 'initiators' and 'promotors', or, less specifically, early-stage and late-stage effectors. If biomarkers (e.g. biochemical markers such as CEA and AFP) were able to identify the cancer stage already reached, then the subsequent risk alteration associated with an exposure of interest would tell something about its stage (and mode?) of action.

Early detection of altered health risks in monitored populations

The above two sections refer to studies of etiology and causal mechanism. In some circumstances, where etiology and causation are not at issue but where there is a need to identify any (suspected) change in population health risk, then biomarkers may be used for monitoring purposes.

This approach is well known in the occupational arena. Biological monitoring of exposed populations can provide early evidence that critical levels of exposure have occurred and that the risk of clinical disease has increased. For example, workers in lead-exposed jobs have long had their blood lead (PbB) concentrations monitored. Lead enters the body from multiple environmental sources (food, water and air), and, in the first instance, is measurable in the blood. Because lead is concentrated within the red blood cells, PbB gives an integrated measure of lead exposure over the preceding several months. It also integrates the exposure measurement across all exposure media.

Increasingly, various occupational exposures are being monitored via cytogenetic studies (Gun & McMichael, 1990). This includes ionizing radiation and solvent exposures. The presumption here is that evidence of macroscopic damage to chromosomes indicates an increased probability of mutational events, which carry an increased risk of cancer.

A mutation known to be particularly strongly associated with a specific type of cancer may be useful in monitoring changes in the risk status of

the population in relation to the future incidence of that cancer. An interesting example arises in relation to the prediction that the ongoing depletion of stratospheric ozone will cause an increase in the incidence of skin cancers (Madronich & de Gruijl, 1994). Epidemiological monitoring of the incidence of skin cancers (basal cell, squamous cell, melanoma) would therefore be desirable in 'sentinel populations' over a range of latitudes. While data on melanoma can be captured by cancer registry data, basal cell and squamous cell cancers may be less reliably reported. However, since these clinically defined end-points may take years to become manifest, earlier predictive biomarkers are needed. One possibility is to use certain dimer-forming mutations of the *p53* gene in skin cells, which appear to be induced specifically by UV exposure (Ziegler *et al.*, 1993; Nakazawa *et al.*, 1994). It has also been suggested that some non-human animals may be particularly useful as indicator species—for instance, the exposed ear epithelium of sheep may be especially sensitive to UV-induced molecular biological damage.

Confirmation of the predictivity of biomarkers in controlled intervention studies

Controlled intervention studies provide an opportunity to test whether the predictive power of the biomarker is the same under the strict conditions of controlled experimentation as has been inferred from natural history observations. Within-individual changes in both the biomarker and the clinical end-point are studied in response to the intervention. Confirmation that an intervention-induced change in the biomarker is followed by a change in the clinical end-point establishes the predictive significance of the biomarker (at least under those circumstances).

The recent results of the Concorde Trial of zidovudine (AZT) treatment in HIV/AIDS underscore the need for caution in assuming that presumed precursor indices of outcome can be used as a substitute measure for the outcome (Concorde Coordinating Committee, 1994). The Concorde Trial found that drug-associated change in the CD4 count—which refers to a change in T-cell profile and is a widely recognized index of damage to the immune system—was predictive of neither the subsequent incidence of clinical AIDS nor survival.

Other considerations

The use of biomarkers, especially within a particular assumed mechanistic framework, may suggest, mistakenly, that the mechanism is universal, i.e. that it applies in all populations. However, the significance—i.e. the predictive power—of biomarkers may vary considerably between populations.

This variation may reflect genetically based differences in disease processes or rates of progression, or it may reflect culturally determined variations in those processes. Most non-acute disease processes have complex etiologies, with webs of causation that differ among geographic, ethnic, genetic and socio-economic groups. For example, the prevalence of chromosomal abnormalities in white blood cells in healthy population samples was found to be considerably greater in West Africa (Gambia) than in Europe, despite the fact that the former population has a much lower lifetime risk of cancer overall (Miele *et al.*, 1996).

Conclusions

The most widespread use of biomarkers in epidemiology has been to improve the quality of exposure assessment. There is also now a growing use of metabolic and genetic biomarkers to identify population subgroups at differing susceptibility to exposure-related disease.

The application of biomarkers as predictors of disease end-points is much less well-developed. This reflects the intrinsic difficulty in obtaining the prospective evidence that would confirm the predictivity of the biomarker, and the uncertainties that persist so long as the causal role of the biomarker (or its underlying phenomena) in the disease process under investigation is unclear.

Beyond their established role in screening for disease and their emerging role in monitoring high-risk populations, there is a need to evaluate critically the use of biomarkers in the enhancement of epidemiological studies of disease etiology and its primary prevention. Biomarkers can sometimes improve the definition of disease entities. They may also, in future, become increasingly useful as early, surrogate, indices of clinical end-points, both for cancer and other diseases. In general, however, much remains to be learnt about their predictive significance before that latter goal can be widely realized.

References

ABC (Alpha-Tocopherol, Beta-Carotene Cancer Prevention) Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New Engl. J. Med.*, 330, 1029–1035

Allen, S.J., Wild, S.P., Wheeler, J.G., Riley, E.M., Montesano, R., Bennett, S., Whittle, H.C., Hall, A.J. & Greenwood, B.M. (1992) Aflatoxin exposure, malaria and hepatitis B infection in rural Gambian children. *Trans. Royal Soc. Tropical Med. Hyg.*, 86, 426–430

Baker, R.J., Van Den Bussche, R.A., Wright, A.J., Wiggins, L.E., Hamilton, M.J., Reat, E.P., Smith, M.H., Lomakin, M.D. & Chesser, R.K. (1996) High levels of genetic change in rodents of Chernobyl. *Nature*, 380, 707–708

Bonassi, S., Abbondandolo, A., Camurri, L., Dal Pra, L., De Ferrari, M., Degrassi, F. & Forni, A. (1995) Are chromosome aberrations in circulating lymphocytes predictive of future cancer onset in humans? Preliminary results of an Italian cohort study. *Cancer Genet. Cytogenet.*, 79, 133–135

Bressac, B., Kew, M., Wands, J. & Ozturk, M. (1991) Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature*, 350, 429–431

Bryant, M.S., Skipper, P.L., Tannenbaum, S.R. & MacLure, M. (1987) Hemoglobin adducts of 4-aminobiphenyl in smokers and nonsmokers. *Cancer Res.*, 47, 612–618

Cartwright, R.A., Glashan, R.W., Rogers, H.J., Ahmad, R.A., Barham-Hall, D., Higgins, E. & Kahn, M.A. (1982) Role of *N*-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer. *Lancet*, 2, 842–845

Concorde Coordinating Committee (1994) Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. *Lancet*, 343, 871–881

Dubrova, Y.E., Nesterov, V.N., Krouchinsky, N.G., Ostapenko, V.A., Neumann, R., Neil, D.L. & Jeffreys, A.J. (1996) Human minisatellite mutation rate after the Chernobyl accident. *Nature*, 380, 683–686

Duthie, S.J., Ross, M. & Collins, A.R. (1995) The influence of smoking and diet on the hypoxanthine phosphoribosyltransferase (hprt) mutant frequency in circulating T lymphocytes from a normal human population. *Mutat. Res.*, 331, 55–64

Duthie, S.J., Ma, A., Ross, M.A. & Collins, A.R. (1996) Antioxidant supplementation decreases oxidative DNA damage in human lymphocytes. *Cancer Res.*, 56, 1291–1295

Fearon, E.R. & Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, 61, 759–767

Gardner, M.J., Snee, M.P., Hall, A.J., Powell, C.A., Downes, S. & Terrell, J. (1990) Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *Br. Med. J.*, 300, 423–429

Gun, R.G. & McMichael, A.J. (1990) The expanding role of biological markers in occupational health research and monitoring. *J. Occup. Health Safety (ANZ)*, 6, 413–417

Hagmar, L., Brogger, A., Hansteen, I.L., Heim, S., Hogstedt, B., Knudsen, L. & Lambert, B. (1994) Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic study group on the health risk of chromosome damage. *Cancer Res.*, 54, 2919–2922

Hanssen, H.P., Agarwal, D.P., Goedde, H.W., Bucher, H., Huland, H., Brachmann, W. & Ovenbeck, R. (1985) Association of N-acetyl-transferase polymorphism and environmental factors with bladder carcinogenesis. *Eur. Urol.*, 11, 263–266

Hattis, D. & Silver, K. (1993) Use of biomarkers in risk assessment. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology: Principles and Practices*, New York, Academic Press, pp. 251–273

Hein, H.O., Sorensen, H., Snadicani, P. & Gyntelberg, F. (1993) Alcohol consumption, Lewis phenotypes, and risk of ischaemic heart disease. *Lancet*, 341, 392–396

Hennekens, C.H., Buring, J.E., Manson, J.E., Stampfer, M., Rosner, B., Cook, N.R., Belanger, C., LaMotte, F., Gaziano, J.M., Ridker, P.M., Willett, W. & Peto, R. (1996) Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N. Engl. J. Med.*, 334, 1145–1149

Hsu, I.C., Metcalf, R.A., Sun, T., Welsh, J.A., Wang, N.J. & Harris, C.C. (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature*, 350, 427–428

Hulka, B.S. (1991) Epidemiological studies using biological markers: issues for epidemiologists. *Cancer Epidemiol. Biomarkers Prev.*, 1, 13–19

Lithell, H.O., McKeigue, P.M., Berglund, L., Mohsen, R., Lithell, U. & Leon, D. (1996) Relationship of birthweight and ponderal index to NIDDM and insulin response to glucose challenge in men aged 50–60 years. *Br. Med., J.*, 312, 406–410

Lower, G.M., Jr, Nilsson, T., Nelson, C.E., Wolf, H., Gamsky, T.E. & Bryan, G.T. (1979) *N*-Acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology: preliminary results in Sweden and Denmark. *Environ. Health Perspect.*, 29, 71–79

McMichael, A.J. (1994) 'Molecular epidemiology': new

pathway or new travelling companion? *Am. J. Epidemiol.*, 140, 1–11

McMichael, A.J., Jensen, O.M., Parkin, D.M. & Zaridze, D.G. (1984) Dietary and endogenous cholesterol and human cancer. *Epidemiol. Rev.*, 6, 192–210

Madronich, S. & de Gruijl, F.R. (1994) Skin cancer and UV radiation. *Nature*, 366, 23

Miele, M., Donato, F., Hall, A.J., Whittle, H., Chapot, B., Bonatti, S., De Ferrari, M., Artuso, M., Gallerani, E., Abbondandolo, A., Montesano, R. & Wild, C.P. (1996) Aflatoxin exposure and cytogenetic alterations in individuals from The Gambia, West Africa. *Mutation Res.*, 349, 209–217

Minchin, R.F., Kadlubar, F.F. & Ilett, K.F. (1993) Role of acetylation in colorectal cancer. *Mutat. Res.*, 290, 35–42

Muñoz, N., Hayashi, M., Lu, J.B., Wahrendorf J, Crespi, M. & Bosch, F.X. (1987) Effect of riboflavin, retinol, and zinc on micronuclei of buccal mucosa and of esophagus: a randomized double-blind intervention study in China. *J. Natl Cancer Inst.*, 79, 687–691

Nakazawa, H., English, D., Randell, P.L., Nakasawa, K., Martel, N., Armstrong, B.K. & Yamasaki, H. (1994) UV and skin cancer specific p53 gene mutation in normal skin as biologically relevant exposure. *Proc. Natl Acad. Sci.*, 91, 360–364

Oda, T., Tsuda, H., Scarpa, A., Sakamoto, M. & Hirohashi, S. (1992) P53 Gene mutation spectrum in hepatocellular carcinoma. *Cancer Res.*, 52, 6358–6364

Ozturk, M., Bressac, B., Puisieux, A. & 26 other collaborators (1991) p53 Mutation in hepatocellular carcinoma after aflatoxin exposure. *Lancet*, 338, 1356–1359

Perera, F. & Santella, R. (1993) Carcinogenesis. In: Schulte, P.A. & Perera, F.P., eds, *Molecular Epidemiology. Principles and Practices*. New York, Academic Press. pp. 277–300

Roberts-Thomson, I., Ryan, P., Khoo, K.K., Hart, W.J., McMichael, A.J. & Butler, R.N. (1996) Diet, acetylator phenotype and risk of colorectal neoplasia. *Lancet*, 347, 1372–1374

Sheu, J.-C., Huang, G.-T., Lee, P-H., Chung, J.-C., Chou, H.-C., Wang, J.-T., Lee, H.-S., Shih, L.-N., Yang, P.-M., Chen, D.-S. & Wang, T.-H. (1992) Mutation of the p53 gene in hepatocellular carcinoma in Taiwan. *Cancer Res.*, 52, 6098–6100

Sidransky, D., Tokino, T., Hamilton, S.R., Kinzler, K.W., Levin, B., Frost, P. & Vogelstein, B. (1992) Identification of ras oncogene mutations in the stool of patients with curable colorectal cancer. *Science*, 256, 102–105

Taylor, J.A., Sandler, D.P., Bloomfield, C.D., Shore, D.L., Ball, E.D., Neubauer, A., McIntyre, O.R. & Liu, E. (1992) Ras oncogene activation and occupational exposures in acute myeloid leukemia. *J. Natl Cancer Inst.*, 84, 1626–1632

Taylor, P.R., Li, B., Dawsey, S.M., Li, J-Y., Yang, C.S., Guo, W. & Blot, W.J., (1994) Linxian Nutrition Intervention Trials Study Group. Prevention of oesophageal cancer: the nutrition intervention trials in Linxian, China. *Cancer Res.* (suppl. 54), 2029s–2031s

Vahakangas, K.H., Samet, J.M., Metcalf, R.A., Welsh, J.A., Bennett, W.P., Lane, D.P. & Harris, C.C. (1992) Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. *Lancet*, 339, 576–580

Vineis, P., Bartsch, H., Caporaso, N., Harrington, A.M., Kadlubar, F.F., Landi, M.T., Malaveille, C., Shields, P.G., Skipper, P., Talaska, G. & Tannebaum, S.R. (1994) Genetically based *N*-acetyltransferase metabolic polymorphism and low-level environmental exposure to carcinogens. *Nature*, 369, 154–156

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M. & Bos, J.L. (1988) Genetic alterations during colorectal tumor development. *N. Engl. J. Med.*, 319, 525–532

Ziegler, A.M., Leffell, D.J. & Kunala, S. (1993) Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proc. Natl Acad. Sci. USA*, 90, 4216–4220

Corresponding author

A.J. McMichael
Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

The use of biomarkers to study pathogenesis and mechanisms of cancer: oesophagus and skin cancer as models

R. Montesano, P. Hainaut and J. Hall

Recent advances in molecular biology have made it possible to use genetic alterations associated with cancer as biomarkers to study the pathogenesis and mechanisms of cancer. However, the lessons that can be drawn from the analysis of alterations in a particular cancer gene are extremely dependent upon the biological context in which they arise.

In this article, we discuss the biological significance of alterations in the *p53* tumour suppressor gene in cancers of the oesophagus and of the skin. In both tissues, different forms of cancer occur at high frequency (squamous-cell carcinoma and adenocarcinoma in the oesophagus; squamous-cell carcinoma, basal-cell carcinoma and melanoma in the skin).

We show that specific patterns of *p53* alteration occur in these various cancers and that analysis of these alterations is useful to make inferences about the etiopathogenesis of cancers of the oesophagus and of the skin.

In the last decade, considerable progress has been made on the identification of the genetic alterations whose accumulation leads to the development of cancer. The genes concerned can be divided into three groups: oncogenes, tumour suppressor genes and genes involved in DNA repair. The genetic alterations found include point mutations, deletions, translocations, allelic losses, amplifications and deregulation of gene expression. The nature of the genetic alterations and the type of gene modified vary among different tumours. Activating *ras* mutations, for instance, frequently occur in pancreatic carcinomas (Berrozpe *et al.*, 1994), but are absent in cancer of the oesophagus (Hollstein *et al.*, 1988). The temporal sequence of occurrence of genetic alterations may also differ between tumour types, with certain changes occurring at an early stage of cancer development in one type of tumour and at a late stage in another type. This is shown schematically in Fig. 1 for colon cancer and glioblastoma.

The relevance of genetic alterations in the etiology and pathogenesis of lung and breast cancers has been discussed in recent reviews (Gadzar *et al.*, 1994; Greenblatt *et al.*, 1994; Hulka & Stark, 1995; Hainaut *et al.*, 1996). Comparison between these different cancers reveals that there is no fixed rule on the nature and sequence of genetic alterations in cancer development. At the same time, it is apparent that some of the genetic alterations, in particular *p53* mutations frequently occur in distinct tumour types and provide valuable insights not only into the etiology but also into the pathogenesis and clinical management (diagnosis and prognosis) of some cancers.

In this respect, the relevance of genetic alterations in oesophageal cancer and in cancer of the skin is discussed in more detail.

Molecular pathogenesis of oesophageal cancer

Oesophageal cancers comprise squamous-cell carcinoma and Barrett's adenocarcinoma, two types of tumour with different etiology and pathogenesis. Squamous-cell carcinoma (SCC) has been associated with exposure to tobacco, betel chewing, alcohol and various dietary components, with a large geographical variation in incidence (Muñoz & Castellsagué, 1994). Adenocarcinomas (ADCs) of the oesophagus develop from Barrett's oesophagus, a condition in which normal squamous epithelium

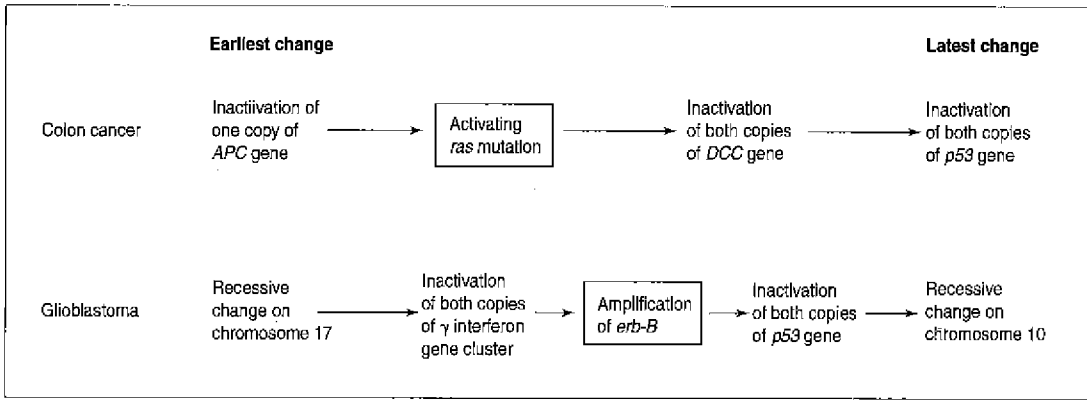

Figure 1. Temporal sequence of genetic events during the progression of colon cancer and of glioblastoma (for explanation, see text); redrawn from Murray & Hunt, 1993

is replaced by metaplastic columnar epithelium. This condition is the result of a prolonged gastro-oesophageal reflux, and these patients have a more than 100-fold higher risk than the general population of developing oesophageal ADC, a cancer that now accounts for approximately 50% of all oesophageal cancers in the USA (Blot *et al.*, 1993).

Prevalence of allelic losses

Table 1 shows the allelic losses in oesophageal cancers (SCC and ADC) that occur with high frequency (>40% of informative cases), based on multiple and reproducible studies. Putative candidate genes localized in the chromosomal areas of allelic losses and the occurrence of mutations in these genes are also indicated (for review see Montesano *et al.*, 1996, and references therein).

No significant differences were detected in the prevalence of allelic losses at various loci in the two types of oesophageal cancer, SCC and ADC. Allelic loss of chromosome 17p occurs very frequently in oesophageal cancer, as in many human solid tumours. In addition, there is a high prevalence of mutations in the *p53* tumour suppressor gene (localized at 17p13) in tumours that retain two 17p alleles (Huang *et al.*, 1993; Maesawa *et al.*, 1994). This suggests that *p53* mutation is an early event in oesophageal carcinogenesis which occurs prior to allelic deletion in 17p.

Allelic loss in oesophageal cancer also occurs with high prevalence in chromosomes 13q, 5q and, possibly, 18q, and in regions of these chromosomes where the *Rb*, *APC/MCC* and *DCC* genes are located. However, *APC/MCC* or *DCC* genes may not be principal targets for loss of heterozygosity (LOH) on 5q and 18q. In particular, mutations in the *APC* gene are very rarely detected (2 out of 163 tumours) (Aoki *et al.*, 1994; Powell *et al.*, 1994; Shibagaki *et al.*, 1994). It is reasonable to assume that other genes with tumour suppressor activity may be located in these chromosomal regions. No mutations were detected in the *Rb* gene in tumours with or without 13q allelic loss, although the Rb protein was found to be absent in a significant fraction of oesophageal SCC.

Another frequent allelic loss in both SCC and ADC involves chromosome 9p21-22, the region in which the gene *MTS-1* is located. This gene encodes the p16 protein, a specific cyclin-dependent kinase (CDK) inhibitor checking G1- to S-phase transition in the cell cycle. This function of p16 is lost when the *MTS-1* gene is mutated. However, it is still unclear whether the *MTS-1* gene alone or other gene(s) present in the same chromosomal region are the target of allelic deletion.

Alterations of cell cycle regulatory proteins

Genetic alterations leading to the constitutive activation of the CDK/cyclin D1 pathway appear to be common in SCC of the oesophagus. These alterations occur at several distinct levels, including (a) inactivation of the *p16/MTS-1* by diverse mechanisms, including deletion and missense mutation; (b) amplification and overexpression of the *cyclin D1* gene (on chromosome 11q13); and (c) deletion

and inactivation of the retinoblastoma susceptibility gene *Rb* (on chromosome 13q14). Amplification and overexpression of the *cyclin D1* gene is the most common of these genetic events (for review, see Montesano *et al.*, 1996).

In a study of 50 oesophageal tumours of varying geographical origin, amplification and overexpression of the *cyclin D1* gene was found in 32% of the tumours, and loss of Rb protein expression in 17. The tumours with *cyclin D1* alterations exibited normal levels of Rb expression, while those that did not express *Rb* did not show amplification or overexpression of *cyclin D1* (Jiang *et al.*, 1993). Similar findings have been observed in other tumour types.

Several studies have described missense mutations of the *p16/MTS-1* gene, but the reported frequencies differ widely. Recent evidence suggests that the prevalence of *p16/MTS-1* mutations is in fact quite low (10%), and that mutations occur in both exon 1 and 2 of the gene (Esteve *et al.*, 1996).

In other tumours, recent evidence suggests that the apparent loss of *p16* activity could also result from aberrant or strongly decreased *p16/MTS-1* gene expression. The *p16/MTS-1* locus has a complex structure and encodes two transcripts with distinct protein coding potential that are differentially regulated during the cell cycle. Deregulation of *p16/MTS-1* activity may thus result from an imbalance between the levels of expression of the two transcripts (Stone *et al.*, 1995). It is also possible that downregulation of the protein rather than point mutation of the gene may be a common mechanism of *p16/MTS-1* inactivation (Sun *et al.*, 1995).

Aberrant expression of *p16/MTS-1, cyclin D1* or *Rb* may have similar functional consequences: inactivation of the suppressor function of Rb and promotion of the activation of E2F proteins by direct or indirect mechanisms (Fig. 2). This model

Table 1. Frequency of allelic loss in oesophageal cancers

	Tumours with allelic loss/ informative tumours (%)				
Chromosome	**SCC**	**ADC**	**Minimal area of loss**	**Candidate genes**	**Mutations**
3p	25/52 (40–64)		3p21.3	*hMLH-1*	
5q	50/94 (36–80)	34/51 (63–75)	5q21.2	*APC/MCC*	Very rare
9p	55/95 (45–65)	8/17 (47)	9p22	*MTS-1, IFNA*	0 <> 40
9q	11/18 (60)		9q31	*ESS1*	
13q	66/134 (41–54)	28/67 (36–43)	13q14.1	*Rb*	No mutations
17p	65/124 (43–65)	47/54 (72–94)	17p13	*p53*	High
17q	56/91 (62)	10/18 (56)	17q21.3 17q11.2-q12	*BCRA1, erbB2,* *CSF-3, NF-1, ITB4*	
18q	26/79 (23–46)	(24)	18q23.3	*DCC*	Rare

SCC, squamous-cell carcinoma; ADC, adenocarcinoma

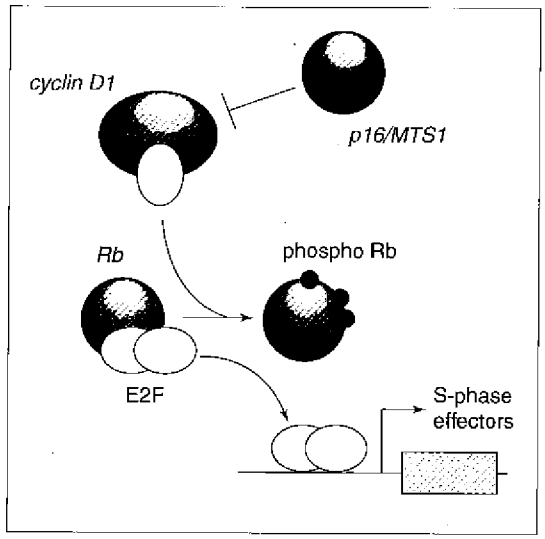

Figure 2. Interplay between the products of the *cyclin D1*, *p16/MTS-1* and retinoblastoma susceptibility genes (*Rb*) in the regulation of cell cycle progression from G1 to S. Cell progression from G1 to S requires, among other things, activation of specific CDKs (CDK4 and CDK6) in association with cyclin D1. The active CDK/cyclin D1 complex phosphorylates sequentially the Rb protein, thus releasing Rb-bound transcription factors of the E2F family. Free E2Fs transactivate genes that are essential for entry into S and DNA replication (S-phase effectors). CDK/cyclin D1 activity is negatively regulated by binding of several cyclin kinases inhibitors, including p16/MTS-1. p16/MTS-1 is thought to be activated in response to growth control stimuli through pathways which are not fully elucidated.

provides a paradigm explaining why tumours with indistinguishable biological characteristics show heterogeneity at the causative genetic level.

The high frequency of genetic aberrations in pathways controlling G1/S transition may reflect a requirement to abrogate physiological G1/S control to allow oesophageal cells to progress towards malignancy. Indeed, the homeostasis of the oesophageal mucosa, as is the case with all surface epithelia, is dependent upon a delicate equilibrium between cell renewal, differentiation and death. Switching off any of the molecular controls balancing these three processes may compromise the normal life cycle of mucosal cells. Escape from growth control, and reduced rates of cell differentiation and cell death are critical steps towards the acquisition of further genetic lesions and functional alterations that lead to the expression of malignant phenotypes.

Mutation spectrum of the p53 gene

Among 240 cases of oesophageal SCC analysed in the literature, 110 (45.8%) have been found to contain mutations that have been confirmed and identified by DNA sequencing. The prevalence of *p53* mutations is even higher for ADC, with 46 cases of mutant *p53* identified in 64 patients screened (71.8%) (see Montesano *et al.*, 1996).

SCC and ADC of the oesophagus differ strikingly in the pattern of mutations in *p53* (Fig. 3). *p53* mutations in ADC show a very high frequency of transition at CpG dinucleotides (~63%). To date, this is the highest level of CpG transition found in any cancer type (other cancer types with frequent CpG transitions are colon carcinoma and pancreatic cancer, with, respectively, ~ 46 and ~36%). G:C to T:A transversions and mutations at A:T base pairs are comparatively rare (taken together, ~14%). In contrast, in SCC, CpG transitions are less frequent than in most other tumour types (~18%), whereas mutations at A:T base pairs account for ~31% of all mutations. The spectrum of *p53* mutations in SCC is thus indicative of the involvement of exogenous carcinogens, in agreement with epidemiological data supporting the role of environmental agents, in particular tobacco, nutritional components and alcohol. The high frequency of mutations at A:T base pairs may reflect enhanced depurination of DNA upon reaction of carcinogens with adenine and/or exposure to DNA-reactive agents such as acetaldehyde, a metabolite of ethanol. The nature of *p53* mutations in ADC is more difficult to interpret. Transitions at CpG dinucleotides are generally considered as the hallmark of mutations occurring spontaneously by hydrolytic deamination of 5-methylcytosine. Recent evidence indicates that the mutability of CpG dinucleotides might also result from enzymatic deamination and methylation by methyltransferases, which bind with high affinity to the premutagenic DNA mismatches G:U and G:T, thereby preventing their efficient repair by mismatch repair proteins (Shen *et al.*, 1995; Yang *et al.*, 1995). Further studies are needed to identify and characterize such mechanisms in human cells and their possible role in the genesis of *p53* mutations in ADC.

In many areas of the world, tobacco and alcohol have been identified as major factors cooperating in the etiopathogenesis of SCC of the oesophagus. Among 91 SCC patients for whom reliable data on

exposure to tobacco and/or alcohol are available, the distribution of *p53* mutations reveals a strong relationship with tobacco smoking (Montesano *et al.*, 1996). Only 20% of non-smoker SCC patients have *p53* mutations, in contrast with 80% in patients who smoked more than 20 cigarettes/day. Even lighter smoking (less than 20 cigarettes/day) increases the *p53* mutation frequency to a value up to 50%. The relationship with alcohol consumption is less clear. However, the relative impact of each risk factor is difficult to assess, as most of the patients with *p53* mutations were exposed to both tobacco and alcohol, and the contribution of other less well-defined risk factors is unknown.

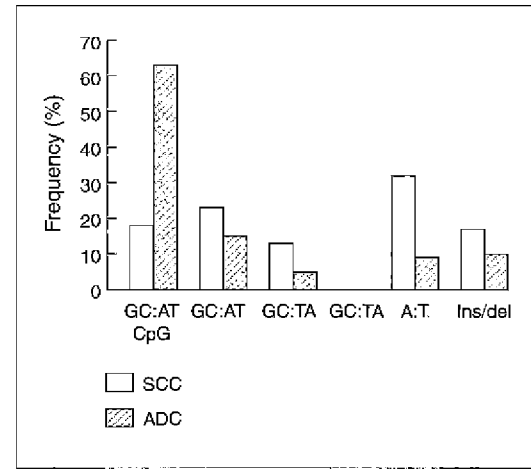

Figure 3. Spectrum of *p53* mutations in squamous-cell carcinoma (SCC) and adenocarcinoma (ADC) of the oesophagus. Data from the *p53* mutation database (Hollstein *et al.*, 1996) are expressed as a percentage of the total number of mutations found in SCC or in ADC.

Molecular pathogenesis of skin cancer

A causal association between UV exposure, defective repair of UV-induced DNA photoproducts and non-melanocytic skin cancer has long been postulated. This cancer is classified into two major histological types: basal-cell carcinoma (BCC) and squamous-cell carcinoma (SCC), the former being the commoner type in white populations. Descriptive studies in whites in North America, Australia and several other countries have also shown a positive association between incidence and mortality from melanoma and residence at lower latitudes. Studies of migrants suggest that the risk of melanoma is related to solar radiant exposure at the place of residence in early life, and the results from case–control studies are generally consistent with positive associations with residence in sunny environments throughout life, in early life and even for short periods in early adult life and with measurements of cumulative sun damage.

UVA, UVB and UVC radiation produce measurable DNA damage in human skin cells *in vivo* at doses commonly experienced by humans. Of the DNA photoproducts, the cyclobutane-type pyrimidine dimers and the pyrimidine-pyrimidone 6-4 photoproducts have been shown to be cytotoxic and mutagenic in human cells, although the relative contribution of each type remains to be fully elucidated. Exposure to UV radiation has been shown to increase the expression of various cellular genes and to have profound effects on the immunological system which may contribute to the development of skin cancer. Evidence for the involvement of DNA photoproducts and their repair in human skin carcinogenesis originally came from the work of Cleaver showing that cells from patients with xeroderma pigmentosum (XP), a cancer-prone inherited disorder, are defective in the excision repair of UV-induced pyrimidine dimers (for review, see Bootsma & Hoeijmakers, 1994).

Alteration of DNA repair and risk of basal- and squamous-cell carcinoma

Mammalian cells have a variety of repair mechanisms that will remove the DNA lesions produced by UV irradiation. The pathway involved in the repair of the commonest UV-induced DNA adducts by UVB and UVC is via the nucleotide excision repair (NER) pathway. The molecular basis of this repair pathway, which eliminates a remarkable number of different, structurally unrelated lesions, involves at least five distinct steps: damage recognition; incision of the damaged strand on both sides of the lesion at some distance from the lesion leaving the non-damaged strand intact; removing the damage-containing oligomer; gap-filling DNA synthesis; and sealing of the DNA by DNA ligase. In human cells, it would appear that two NER subpathways are operating, one dealing with the rapid and efficient removal of lesions blocking ongoing transcription (transcription-coupled repair), and the second resulting in the slower and less-efficient removal of bulk DNA including the non-transcribed strand of active genes (genome overall repair).

Individuals lacking components of the NER pathway have been found to have increased sun sensitivity and, in some cases, increased risk of skin cancer. This is observed in XP patients who show sun sensitivity and increased cancer risk in addition to cutaneous manifestations, including neurological degeneration and pigmentation abnormalities.

Within the XP phenotype, large variations in repair activity are noted (ranging from 2 to 80% of 'normal' levels). If XP is considered to represent the lower range of repair capabilities, those individuals expressing a reduced repair response may be at an increased risk of cancer. It could be imagined that, due to the multiprotein complexes involved in the NER pathway, minor alterations in protein structure might produce dramatic alterations in protein–protein interactions and subsequent alterations in the DNA-repair capacity (DRC). Due to the multistep nature of the NER pathway, an assay that reflects the overall DRC and that can be carried out on easily available biological material, such as lymphocytes, would be ideal for population-based studies to assess the role of DRC in modulating cancer risk. Such an assay was developed by Grossman and his collaborators (Athas *et al.*, 1990) and has subsequently been used in two published studies designed to evaluate the role of DRC as a risk factor for non-melanocytic skin cancer. In this assay, blood lymphocytes from subjects are cultured and transfected with either control or UV-irradiated plasmids containing a reporter gene (the chloramphenicol-acetyltransferase gene *CAT*), and the repair capacity is determined by measuring *CAT* gene expression in protein extracts prepared from transfected cells.

Grossman and his collaborators applied the assay in a study population in Baltimore, Maryland. This population was made up of Caucasians (aged 20–60 years) who had lived in Baltimore or its suburbs for most of their lives. The DRC was measured in 88 patients with a history of one or more histologically confirmed BCCs, and in 135 control subjects. The mean *CAT* activity with plasmids UV-irradiated at 700 J/m^2 was 8.00% (SD = 2.2) in controls without family history of actinic keratosis, 7.28% (SD = 2.2) in controls with a family history of BCC or a previous history of actinic keratosis, and 7.35% (SD = 2.0) in cancer cases ($P = 0.103$; however, when cancer cases were compared with subjects with family history, the P-value was 0.047). They also found that DRC declined with age in both cases and controls (Wei *et al.*, 1993).

In a second study by Hall *et al.* (1994), the DRC was measured in subjects involved in a population-based study of the incidence and prevalence of non-melanocytic skin cancer in Geraldton, Western Australia. Subjects with SCC and BCC were considered separately, and for each case one control was chosen, matched by age and sex. The mean levels of *CAT* activity in cells transfected with plasmids irradiated at 700 J/m^2 were found to be generally higher in cases than in controls, and there was little difference between cases and controls in the DRC capacity according to the UV. The DRC showed little association with age, sex and viability of the lymphocytes, but was positively correlated with blastogenic rate ($P = 0.055$). Thus, no evidence was found that subjects with non-melanocytic skin cancer had lower repair capacities than the controls, nor was any statistically significant difference in repair activity detected between the two groups. Only in the young subjects with SCC was there any suggestion that cases had lower repair capacities, although this difference could be due to chance ($P = 0.53$).

The differences between the results of these two studies may be explained by several factors. Firstly, the age range of the subjects in the Baltimore study was larger than that in the Australian study, with 21.6% of the BCC cases under the age of 40 at the time of sample collection. As the apparent difference in DRC was most marked in early-onset cases, the possibility of detecting a difference between cases and controls might be reduced due to the older age of the Australian population. A second major difference was in the mean *CAT* activity measured in the assay, which was substantially higher in the Australian study (13% in BCC cases, SD = 6.2%; 12.2% in SCC cases, SD = 7.1%; and 12% in controls, SD = 5.6%). This may indicate a greater degree of random error in the laboratory measurements. The methods used in the studies were not identical, mainly for logistic reasons. The handling of the blood samples is one example. In the Australian study, lymphocyte purification could only be carried out after overnight transport of the samples, whereas Wei *et al.* (1993) processed the samples locally on the same day. Since subsequent cell viability was lower in the Australian study, the pos-

sibility that the surviving cells were not truly representative of the individual cannot be excluded.

Another possible explanation relates to the higher ambient level of solar radiation in Western Australia. Exposure to UV radiation has been shown to increase the expression of various genes. Short-term effects of UV on the ability to repair radiation-induced DNA damage have been noted, but little is known about the long-term effects. It is also possible that the result of high environmental exposures to UV radiation is to mask the DRC differences between cases and controls observed by Wei *et al.* (1993). Recent epidemiological evidence suggests that beyond a certain sun exposure, the risk of BCC does not increase further and that a particular amount of sun exposure delivered in infrequent, probably intense, increments will increase the risk of BCC. This finding may support the possibility that in cases of extreme sun exposure, DRC is no longer a major risk factor for BCC.

In summary, the use of the measurement of DRC in cells to identify individuals at increased risk of UV carcinogenesis and to improve the estimation of the effects of sun exposure is appealing. The assay developed by Grossman and colleagues is potentially applicable to molecular epidemiology. However, its use in such studies has, to date, produced conflicting data, which may reflect differences in experimental approach as well as differences between the populations studied.

There is limited evidence, apart from that coming from XP patients and that obtained by Wei *et al.* (1993), that DRC is altered in patients with skin cancer. A number of studies have examined the ability to repair pyrimidine dimers induced by single exposure to stimulated solar radiation, although many of these studies are limited by small populations (IARC, 1992). The impact of reduced DRC as a risk factor for BCC is potentially significant, especially when the occurrence of increased ambient UV irradiation, due to alterations in ozone levels, is taken into consideration. However, in order to fully assess the contribution of DRC to skin cancer susceptibility, additional studies need to be completed in which the effects of age, family history and sun exposure can be evaluated and linked to other important factors in the carcinogenic process, such as the mutagenic alterations in cellular DNA.

Prevalence of allelic loss

The pattern of chromosome loss has been studied in skin tumours from patients with BCC, SCC, melanoma and Ferguson–Smith syndrome (a syndrome in which the lesions are histologically indistinguishable from SCCs) using microsatellite markers. A high frequency of chromosome 9q loss (up to 68% in informative tumours) has been observed in BCC, within a region where both the naevoid BCC syndrome and Ferguson–Smith syndrome have been mapped (9q22.3–q31). This suggests that the two conditions may reflect mutations in the same gene. A relatively high frequency of LOH (14%) was also observed for chromosome 1q, suggesting that this region may play a role in the development of BCC.

In SCC, loss of chromosome 9 has been found in up to 33% of tumours (16/49). However, the pattern of chromosomal alterations in SCC is different from that observed in BCC, with LOH at 9p (41%), 13q (46%), 17p (33%) and 3p (23%) (Quinn *et al.*, 1994). In primary melanomas, LOH at 9p was the most frequent (47% of informative cases), and it has been suggested that a melanoma susceptibility or tumour suppressor gene may be localized at 9p21-23, a region that contains the cyclin-dependent kinase inhibitor gene *p16/MTS-1*. LOH on chromosome arms 3p, 6q, 10q, 11q and 17p was also relatively frequent. LOH at 3p and 10q was found in lesions of 0.5 mm in depth or less, while LOH at 6q, 11q and 17p was only detected in more invasive tumours. These results suggest that loss of chromosome 9p occurs before loss of other chromosome arms in sporadic cutaneous melanoma.

Activation of oncogenes

A high variability in *ras* gene mutations has been reported in melanomas (from 5 to 27%) and may be explained by the results of Jafari *et al.* (1995). They demonstrated that the predominate mutation was at codon 61 of the N-*ras* gene and exclusively in modular malignant melanomas (31% mutation frequency). No *ras* mutations were reported in superficial spreading melanomas or in lentigo malignant melanomas, demonstrating the need for consideration of the specific pathology when carrying out this type of analysis. Whether *ras* gene mutations play an important role in non-melanocytic skin cancer remains to be fully established. Spencer *et al.* (1995) have found activating

mutations at codon 12 of K-*ras* and at codons 12, 13 and 61 of the H-*ras* gene in 16% of actinic keratoses and in 12% of SCCs. Daya-Grosjean *et al.* (1993) have shown that there is a twofold increase in mutation frequency of the *ras* genes in XP tumours compared with control tumours. The majority of these mutations were at codon 12 in all three *ras* genes, and were located opposite dipyrimidine sequences. This is consistent with a fundamental role of unrepaired DNA damage as an initiating lesion in skin carcinogenesis.

Mutation spectrum of p53 gene

With the exception of cutaneous melanoma, skin cancer is often associated with *p53* mutations. Recent studies have estimated the incidence of *p53* mutations in SCC and BCC to be between 12 and 58%, and the analysis of the mutations has revealed the presence of 'UV- specific' mutations targeted to dipyrimidine sequences, with the majority being C to T transitions. Ziegler *et al.* (1994) have demonstrated that UV-induced *p53* mutations are already detectable in actinic keratosis, with mutations being detected in 60% of 45 cases of actinic keratosis examined from 24 patients. The base changes found implicated sunlight as the mutagen: 89% occurred at adjacent pyrimidines and most were C to T substitutions or CC to TT double base changes.

Immunological response

Over the past few years, it has been realized that exposing the skin to sunlight has profound effects

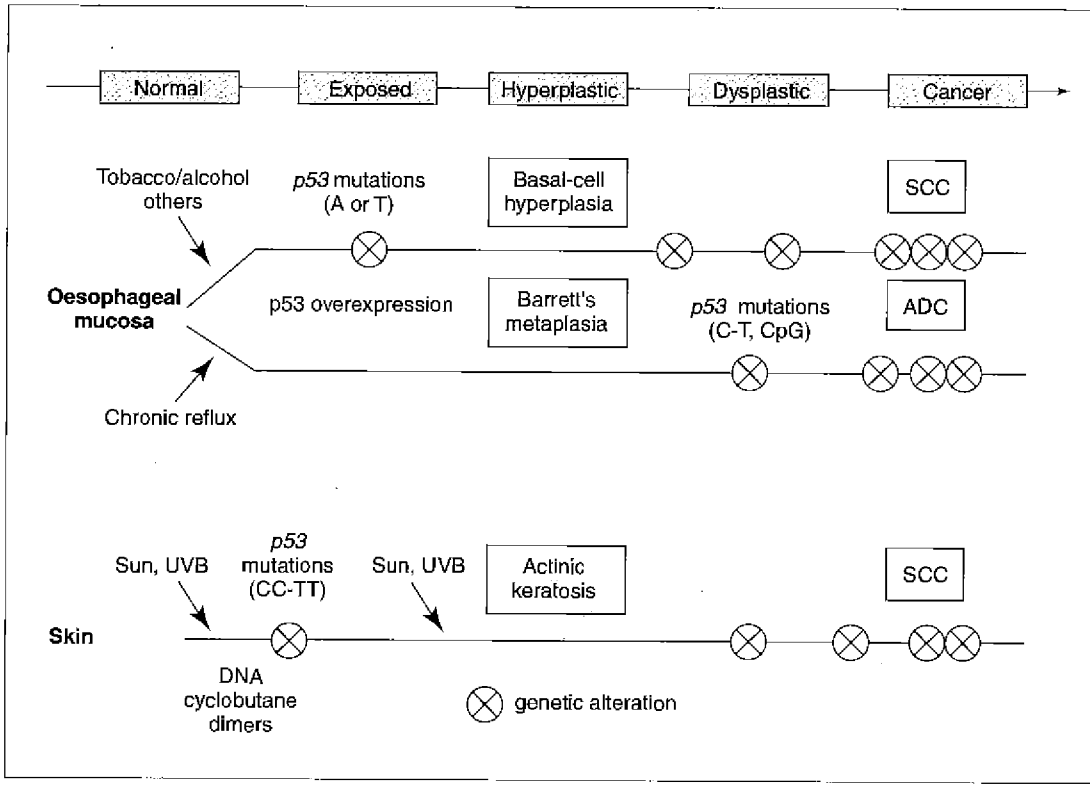

Figure 4. *p53* mutations in the temporal sequences of events leading to cancer of oesophagus [squamous-cell carcinoma (SCC) and adenocarcinoma (ADC)] or of the skin. The progression of the disease is represented from left (normal tissue) to right (cancer). The occurrence of genetic lesions is shown with respect to specific pathological stages. This model illustrates that *p53* alterations occur at different stages and in different contexts in both cancer of the oesophagus and cancer of the skin. In SCC of the oesophagus, mutation of *p53* is thought to be a very early event that occurs in exposed tissues prior to the formation of a dysplastic lesion. In contrast, in ADC, *p53* mutations occur later in the development of the lesion, probably in late dysplasia. In SCC of the skin, mutations occur in normal cells, but the propagation of a clone of cells carrying a *p53* mutation is dependent upon continuous exposure to UV.

on the immunological system and that these immunological changes can contribute to the development of skin cancer and alter resistance to infectious diseases. Studies on the nature and mechanism of the immunological alterations following exposure to UV radiation suggest that UV-induced DNA damage triggers a cascade of events leading to a state of antigen-specific, systemic T-lymphocyte-mediated immunosuppression. It is proposed that with low doses of UV radiation, antigen-presenting cells in the irradiated skin are modified, either directly by the action of UV or indirectly as a consequence of an influx of inflammatory cells or by the local release of cytokines. Larger doses of UV may bring about the release of cytokines systemically, causing an alteration in the activity of antigen-presenting cells. Cytokine imbalances in individuals unable to repair UV-induced DNA damage may explain some of the additional clinical symptoms observed in XP and other sun-sensitive syndromes.

Conclusion

Temporal sequence of genetic alterations in the progression of cancer

Figure 4 summarizes the role of *p53* mutation in the temporal sequence of events leading to cancers of the oesophagus or to cancer of the skin.

Cancer of the oesophagus. In SCC, *p53* overexpression or accumulation of the p53 protein due to prolonged half-life has been detected in normal, non-dysplastic oesophageal cells and, in particular, in the nuclei of proliferative cells in the basal layer of the mucosa. In a number of cases, p53 mutations are already present in the p53-positive cells in normal mucosa. Moreover, in a given patient, multiple and distinct patches of p53-positive cells with different mutations can be detected. This indicates that multiple independent pre-neoplastic foci or tumours develop in the oesophageal squamous mucosa, and it is of interest to elucidate which of these foci have selective growth advantage in cancer development. During tumour development, frequent additional genetic alterations include 3q21.3 and 9q31 allelic losses.

In ADC, overexpression of p53 protein and increased proportions of cells with a G0/G1 DNA content are detected in metaplastic Barrett's epithelium, possibly as a result of DNA damage induced by chronic gastric reflux. Subsequent inactivation of *p53* by mutation and/or allelic result is an abrogation of cell cycle control at the G1/S transition. Consequently, subpopulations of aneuploid cells frequently develop during the later stages of carcinogenesis, with increased proportion of cells in the S-phase and G2-phase. Mutations in the *p53* gene are more common in ADC than in SCC, but are not found in non-dysplastic Barrett's metaplasia, suggesting that they occur later at the transition from high-grade dysplasia to carcinoma.

These findings at the cellular and molecular levels underscore the different etiology and pathogenesis of SCC compared with ADC, and suggest that the genetic alterations observed may represent molecular fingerprints of critical risk factors involved in the development of these two cancers.

Cancer of the skin. The model for human skin carcinogenesis presents some uncertainties as far as the progression of genetic events is concerned. A role for *p53* has been clearly demonstrated by Ziegler *et al.* (1994). *p53* mutations are found in actinic keratosis, and inactivation of *p53* in mouse skin reduces the appearance of sunburn cells, apoptotic keratinocytes generated by overexposure to UV. Skin thus appears to have a *p53*-dependent mechanism which removes 'precancerous' UV-damaged cells via apoptosis. If this response is reduced or altered by *p53* mutation, sunburn might act as a selection pressure favouring the survival of mutated cells. Whereas UV-damaged normal cells will die as sunburn cells, a proportion of the damaged cells with mutant *p53* will be resistant to apoptosis. Continued sun exposure, which produces alterations in immunological surveillance and gene expression, can act as a tumour promoter and results in the clonal expansion of mutated cells to form an actinic keratosis which usually regresses in the absence of sun exposure but progressively enlarges if exposure continues.

Importance of biological context in interpreting the significance of biomarkers

The two examples discussed here indicate that the detailed analysis of specific biomarkers may reveal important information for the understanding of both the etiology and pathogenesis of human cancer. In particular, the analysis of *p53* gene mutation and protein expression is useful to make

inferences about the nature of the etiologic agents involved in these cancers. In addition, in both oesophagus and skin cancers, alteration of *p53* appears to represent a central event in tumour progression. However, the interpretation of the significance of *p53* alteration as a biomarker is dependent upon the biological context. In cancers of the oesophagus, the main consequence of *p53* alterations is likely to be the disruption of G1/S cell cycle control, giving rise to populations of actively growing cells at high risk of becoming precursors to neoplastic lesions. In skin cancers, loss of *p53* function may abrogate a normal mechanism which removes UV-altered cells via apoptosis, thus facilitating the survival of cells containing UV-damaged DNA. Clonal expansion of mutated cells depends upon continuous exposure to the sun until further genetic alterations give rise to squamous cell carcinoma. These two examples emphasize that comparable genetic events may have different impacts depending upon the tissue, cellular and molecular context in which they arise.

Acknowledgements

We gratefully acknowledge the financial support received from the Commission of the European Communities (grant nos EV5V-CT92-0199 and CT 94-0581). We thank Mrs Magali Maillol for typing the manuscript.

References

Aoki, T., Mori, T., Du, X., Nisihira, T., Matsubara, T. & Nakamura, Y. (1994) Allelotype study of esophageal carcinoma. *Genes Chromosom. Cancer*, 10, 177–182

Athas, W.F., Hedayati, M.A., Matanoski, G.M., Farmer, E.R. & Grossman, L. (1990) Developement and field-test validation of an assay for DNA repair in circulating lymphocytes. *Cancer Res.*, 51, 5786–5793

Berrozpe, G., Schaeffer, J., Peinado, M.A., Real, F.X. & Perucho, M. (1994) Comparative analysis of mutations in the p53 and *K-ras* genes in pancreatic cancer. *Int. J. Cancer*, 58, 185–91

Blot, W.J., Devesa, S.S. & Fraumeni, J.F. (1993) Continuing climb in rates of esophageal adenocarcinoma: an update [letter]. *J. Am. Med. Assoc.*, 270, 1320

Bootsma, D. & Hoeljmakers, J.H.J. (1994) The molecular basis of nucleotide excision repair syndromes. *Mutation Res.*, 307, 15–23

Daya-Grosjean, L., Robert, C., Drougard, C., Suarez, H. & Sarasin, A. (1993) High mutation frequency in *ras* genes of skin tumors isolated from DNA repair deficient xeroderma pigmentosum patients. *Cancer Res.*, 53, 1625–1629

Esteve, A., Martel-Planche, G., Sylla, B.S., Hollstein, M., Hainaut, P. & Montesano, R. (1996) Low frequency of *p16*/CDKN2 gene mutations in esophageal carcinomas. *Int. J. Cancer*, 66, 301–304

Gadzar, A.F., Bader, S., Hung, J., Kishimoto, Y., Sekido, Y., Sugio, K., Virmani, A., Fleming, J., Carbone, D.P. & Minna, J.D. (1994) Molecular genetic changes found in human lung cancer and its precursor lesions. In: *The Molecular Genetics of Cancer*, Vol. 59, Cold Spring Harbor, USA, Cold Spring Harbor Laboratory Press, pp. 565–572

Greenblatt, M.S., Bennett, W.P., Hollstein, M. & Harris, C.C. (1994) Mutations in the p53 tumour suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.*, 54, 4855–4878

Hall, J., English, D.R., Artuso, M., Armstrong, B.K. & Winter, M. (1994) DNA repair capacity as a risk factor for non-melanocytic skin cancer—a molecular epidemiological study. *Int. J. Cancer*, 58, 179–184

Hollstein, M.C., Smits, A.M., Galiana, C., Yamasaki, H., Bos, J.L., Mandard, A., Partensky, C. & Montesano, R. (1988) Amplification of EGF receptor gene but no evidence of *ras* mutations in primary esophageal cancers. *Cancer Res.*, 48, 5119–5123

Hollstein, M., Shomer, B., Greenblatt, M., Soussi, T., Hovig, E., Montesano, R. & Harris, C.C. (1996) Somatic point mutations in the *p53* gene of human tumors and cell lines: updated compilation. *Nucleic Acids Res.*, 24, 141–146

Huang, Y., Meltzer, S.J., Yin, J., Tong, Y., Chang, E.H., Srivastava, S., McDaniel, T., Boynton, R.F. & Zou, Z.Q. (1993) Altered messenger RNA and unique mutational profiles of p53 and Rb in human esophageal carcinomas. *Cancer Res.*, 53, 1889–1894

Hulka, B.S. & Stark, A.T. (1995) Breast cancer: cause and prevention. *Lancet*, 346, 883–87

IARC (1992) *IARC Monograph on the Evaluation of Carcinogenoc Risks to Humans: Solar and Ultraviolet Radiation*, Vol. 55, Lyon, International Agency for Research on Cancer

Jafari, M., Papp, T., Kirchner, S., Diener, U., Henschler, D., Burg, G. & Schiffmannn, D. (1995) Analysis of *ras* mutations in human melanocytic lesions: activation of the *ras* gene seems to be associated with the nodular type of human malignant melanoma. *J. Cancer Res. Clin. Oncol.*, 121, 23–30

Jiang, W., Zhang, Y.J., Kahn, S.M., Hollstein, M.C., Santella, R.M., Lu, S.H., Harris, C.C., Montesano, R. & Weinstein, I.B. (1993) Altered expression of the cyclin

D1 and retinoblastoma genes in human esophageal cancer. *Proc. Natl Acad. Sci., USA.*, 90, 9026–9030

Maesawa, C., Tamura, G., Suzuki, Y., Ogasawara, S., Ishida, K., Saito, K. & Satodate, R. (1994) Aberrations of tumor-suppressor genes (p53, apc, mcc and Rb) in esophageal squamous-cell carcinoma. *Int. J. Cancer.*, 57, 21–25

Montesano, R., Hollstein, M. & Hainaut, P. (1996) Genetic alterations in oesophageal cancer and its relevance to etiology and pathogenesis: a review. *Int. J. Cancer Pred. Oncol.*, 69, 225–235

Muñoz, N. & Castellsagué, X. (1994) Epidemiology of oesophageal cancer. *Eur. J. Gastroenterol. Hepatol.*, 6, 649–655

Murray, A. & Hunt, T. (1993) *The Cell Cycle*, Oxford, Oxford University Press

Powell, S.M., Papadopoulos, N., Kinzler, K.W., Smokinski, K.N. & Meltzer, S.J. (1994) APC gene mutations in the mutation cluster region are rare in esophageal cancers. *Gastroenterology*, 107, 1759–1763

Quinn, A.G., Sikkink, S. & Rees, J.L. (1994) Basal cell carcinoma and squamous cell carcinoma of human skin show distinct patterns of chromosome loss. *Cancer Res.*, 54, 4756–4759

Shen, J.C., Zingg, J.M., Yang, A.S., Schmutte, C. & Jones, P.A. (1995) A mutant Hpall methyltransferase functions as a mutator enzyme. *Nucleic Acids Res.*, 23, 4275–4282

Shibagaki, I., Shimada, Y., Wagata, T., Ikenaga, M., Imamura, M. & Ishizaki, K. (1994) Allelotype analysis of esophageal squamous cell carcinoma. *Cancer Res.*, 54, 2996–3000

Spencer, J.M., Kahn, S.M., Jiang, W., DeLeo, V.A. & Weinstein, I.B. (1995) Activated *ras* genes occur in human actinic keratoses, premalignant precursors to squamous cell carcinomas. *Arch. Dermatol.*, 131, 796–800

Stone, S., Jiang, P., Dayananth, P., Tavtigian, S.V., Katcher, H., Parry, D., Peters, G. & Kamb, A. (1995) Complex structure and regulation of the P16 (MTS1) locus. *Cancer Res.*, 55, 2988–2994

Sun, Y., Hildesheim, A., Lanier, A.E., Cao, Y., Yao, K.T., Raab, Traub, N. & Yang, C.S. (1995) No point mutation but decreased expression of the p16/MTS1 tumor suppressor gene in nasopharyngeal carcinomas. *Oncogene*, 10, 785–788

Wei, Q., Matanoski, G.M., Farmer, E.R., Hedayati, M.A. & Grossman, L. (1993) DNA repair and aging in basal cell carcinoma: a molecular epidemiology study. *Proc. Natl Acad. Sci.*, 90, 1614–1618

Yang, A.S., Shen, J.C., Zingg, J.M., Mi, S. & Jones, P.A. (1995) HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res.*, 23, 1380–1387

Ziegler, AM., Jonason, A.S., Leffell, D.J., Simon, J.A., Sharma, H.W., Kimmelman, J., Remington, L., Jacks, T. & Brash, D.E. (1994) Sunburn and skin cancer in the onset of skin cancer. *Nature*, 372, 773–775

Corresponding author

R. Montesano
Unit of Mechanisms of Carcinogenesis,
International Agency for Research on Cancer,
150 cours Albert Thomas, 69372 Lyon Cedex 08,
France

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Comparing measurements of biomarkers with other measurements of exposure

R. Saracci

The issue of the relative merit of biomarkers and alternative measures of exposure arises most commonly in the context of epidemiological studies aimed at hazard detection and quantification. When exposures are from biological agents, biomarkers are usually the first and often the only justifiable choice. In general, however, the relative merit of different types of exposure measurements need to be evaluated on a case-by-case basis. Biomarkers may be affected by random errors, time-related sampling errors, physiological confounding and disease-induced differential error, all of which need to be explicitly evaluated before embarking on the use of a biomarker in a full-scale epidemiological study. Random errors affecting biomarkers may be reduced by replication or combination of measurements, or both. Alternative measurements of exposure can be evaluated against a biomarker when there is adequate evidence for regarding the marker as the true measure of a biologically relevant exposure.

Biomarker measurements: for what purpose?

The term 'exposure' can have, and in the past has had, diverse meanings, ranging from very narrow to very broad: in epidemiology it has become customary to regard an exposure as 'any of a subject's attributes or any agent with which he or she may come in contact that may be relevant to his or her health' (Armstrong *et al.*, 1992, p. 4). This all-embracing definition incorporates variables external to the human body (e.g. concentration of an agent in air, social class, a drug), as well as variables internal to the organism (e.g. sex, blood level of a hormone, a gene), the key feature common to both kinds of variable being that they are considered as potential causes of biological and health effects. In the area of exposure measurement for epidemiological research, the fastest growing sector is represented by biomarker measurements, underpinned by current developments in biology, particularly at cellular, subcellular and molecular levels. In principle, measurements of biomarkers using up-to-date techniques can potentially favourably replace many other types of measurement, in particular direct (physical, chemical or biological) or indirect (questionnaire and record) measurements of environmental variables; questionnaire measurements of biological variables; and measurements of various kinds extractable from existing records.

When comparing the relative merits of biomarker measurements and other types of measurement, the following four purposes for which biomarker measurements are performed should be considered:

Control of exposure level

Biomonitoring is a long established practice, essentially within the industrial setting, which acts as a complement or an alternative to environmental measurements. In this application, a biomarker is regarded (based on evidence or assumptions, or both) as a biologically relevant effect of external agents, and research usually centres on the technology required to achieve the desired level of control of the agent in the environment. To the extent, however, that the biological relevance of the marker is not yet adequately established—at least for the specific purpose of exposure monitoring—the same issues arise as are considered in the next paragraph.

Hazard detection and risk quantification.

For this research purpose, a genuine question arises as to whether a biomarker or another type of mea-

surement is preferable. A biomarker measurement may prove superior in six respects: (1) *individualization*, as the measurement is now performed on the subjects in the study or on biological specimens taken from them; (2) *objectivity*, to the extent that the measurement is independent of the observed subjects' perceptions and substantially independent of the observer if instrumental or laboratory methods are used; (3) *quantitative specificity* to the exposure of interest, to the extent that the method of measurement responds only to the agent of interest with no, or a minimal, response to other agents; (4) *quantitative sensitivity*, to the extent that the method of measurement responds in a quantitative way to the agent down to a very low limit of detection (meaning by 'very low' a concentration well below the minimum likely to induce a detectable biological effect); (5) *biological significance*, to the extent that the biomarker measurement may not only measure exposure but also contribute at least some information on the mechanisms through which an agent may induce a disease, adding plausibility to an observed exposure–response relationship; (6) *cost*, to the extent that biomarker measurements, especially when performed with automated systems capable of combining high accuracy and precision with low cost per unit measurement, can be less expensive than alternatives. The worth of a biological marker measurement in these respects, taken together, may make it preferable to the alternatives, in that it is better capable (all other things being equal) of providing a higher estimate of the risk associated with an exposure and a steeper exposure–response relationship; or, while providing similar estimates, it is less costly. This, however, should never be taken for granted: while for individualization a biomarker measurement performed concurrently within a study is always superior, for objectivity, specificity, sensitivity, biological significance and cost, the relative merits of the biomarker measurement and of an alternative measurement can only be evaluated case by case.

Pathogenesis studies.
These studies focus not on the 'black box' relationship between the primary causal factors, environmental or hereditary (inputs), and disease (output), but on the interrelated sequences of biological events which lead from such etiological factors to disease—obviously for these studies there are no alternatives to the use of biomarkers measurements. A key consideration, however, is that epidemiological studies can only be—in general and exceptions apart—ancillary tools for pathogenesis studies, not because of a lack of pertinent biomarkers but because of their observational nature within the complex settings of human life. Mechanistic insights are gained, for instance, by analysing in human tissue specimens the genetic changes characteristic of the different histological conditions of the colonic mucosa (normal tissue, adenoma, carcinoma). However, to elucidate and demonstrate the causal succession and interconnection of genetic and non-genetic events requires exploration and testing in systems accessible to direct experiment within simplified and controlled settings, e.g. animal models like transgenic and knockout mice. In the same way that descriptive epidemiology data provide etiological insights and suggestions to be fully explored and tested in analytical epidemiological studies, epidemiology in general can provide mechanistic insights and suggestions to be fully analysed and established in experimental studies.

Prognostic studies.
In prognostic studies of outcome, the population investigated is composed of patients rather than 'healthy' people. Again, as a general rule, biomarker measurements dominate the field and questions of alternatives seldom arise. In addition to measurements of biomarkers, however, other kinds of measurement may address different issues and may be necessary as well; for example, the measurement of clinical performance status by a questionnaire in addition to biomarkers such as the histological type of a tumour or the amplification of an oncogene.

These four considerations highlight one point: in essence, it is in the area of hazard detection and risk quantification that issues of merit comparison between biomarkers measurements and other types of measurements arise. The comparison can be cast in the systematic terms of evaluating the properties of measurements methods, in particular their error elements, as done both for categorical and continuous variables by White (this volume), or can be centred on signalling potential pitfalls and strengths in the application of biomarkers

measurements compared to alternatives, as in the following seven sections of this article.

Random error

Consider a 'true' situation, perfectly known without errors, involving two dichotomous exposure variables: lifetime smoking habit, evaluated by questionnaire as S+ and S–, and a marker of biologically relevant exposure (BS+ or BS–) reflecting the sum of all biologically effective lung carcinogens in tobacco smoke, as all pathogenetic paths of tobacco lung carcinogens transit through it. Three features of what can happen in these circumstances (Table 1) are worth noticing:

1. Subjects positive for the marker have a risk of lung cancer 11 times higher than subjects negative for the marker.
2. In the absence of information on the marker, knowledge of smoking status as obtainable by questionnaire gives instead a lower relative risk for lung cancer (7.4). This is because some 2.4% of the BS+ subjects are positive because of general environmental exposures to lung carcinogens, but are lifetime non-smokers; and, conversely, in 20% of smokers, the marker remains negative, for physiological (metabolic) conditions.
3. Once the biomarker status of a subject is known, no further information on risk is added by the knowledge of smoking status: the relative risk for lung cancer conditional on BS+ or BS– status is 1 in both strata.

In Table 2 a random error has been introduced in the biomarker measurement, non-differential in respect to smoking status, reducing both the sensitivity and the specificity of the biomarker to 95% (from 100% in the case of the perfect measurements of Table 1). However, the lifetime smoking status as assessed by the questionnaire measurement is 'robust' under a variety of circumstances and assumed to remain error-free. Table 2 shows that:

1. Measuring exposure via the biomarker and via the questionnaire are now nearly equivalent, the former yielding a relative risk of 8 and the latter yielding a relative risk of 7.4.
2. Once the biomarker status is known, assessing smoking status by questionnaire yields additional information as indicated by the fact that, after stratification for BS, the relative risk for smoking is now 2.4 (reciprocally determining BS would add to an initial knowledge of smoking status).
3. This extra information has, however, a disturbing aspect: if controlling for a variable like BS, intermediate in the sequence between external exposure to tobacco smoke and lung cancer, does not account for all of the effect of tobacco smoke, the common interpretation would be that such an effect is also mediated by pathogenic paths other than the one going through BS. In fact, this is not the case. The appearance of an alternative path is simply due to an error in the measurement of BS. Hence, unless we know the size of this error, we may be unable to decide which of the two interpretations is likely to be correct. Actual measurements of any variable are unavoidably affected by random errors.

Table 1. Lung cancer occurrence in subjects classified by two variables (BS and S) with no error of measurement (sensitivity = specificity = 100%)

	BS+ (n = 820)		BS– (n = 1180)	
	S+ (n = 800)	S– (n = 20)	S+ (n = 200)	S– (n = 980)
Lung cancer	80	2	2	10
No lung cancer	720	18	198	970
	Relative risk (S+/S–) = 1.0		Relative risk (S+/S–) = 1.0	
	Relative risk (BS+/BS–) = 11.0			
	Relative risk (S+/S–) = 7.4			

Table 2. Lung cancer occurrence in subjects classified according to BS (sensitivity= specificity = 95%) and S (sensitivity = specificity = 100%)

	BS+ (n' = 838)		BS– (n' = 1162)	
	S+ (n' = 770)	S– (n' = 68)	S+ (n' = 230)	S– (n' = 932)
Lung cancer	76	3	6	9
No lung cancer	694	65	224	923
	Relative risk (S+/S–) = 2.4		Relative risk (S+/S–) = 2.4	

Relative risk (BS+/BS–) = 8.0
Relative risk (S+/S–) = 7.4

It follows from the simple example just presented that the relative sizes of the errors of the measured variables will determine which pattern of relationship is observable; and in turn the interpretation of the pattern will depend to a large extent on knowledge of the magnitude of these errors. The issue is not restricted to the comparison of biomarkers with alternative measures of exposure, but bears, much more generally, on the possibility of correctly disentangling in an observational study the net effects of individual variables, intercorrelated and measured with different degrees of precision (Phillips & Davey Smith, 1991).

The error of measurement has been expressed here through the sensitivity and the specificity in categorizing the subjects of a population through the measurement of a dichotomous variable. More often, the error will be expressed as one component of the total variance of the measurements of a continuous variable in a population of subjects. Among laboratory analysts, it has instead been (and still is) customary to express the analytical error of measurement in the form of the proportional standard deviation (coefficient of variation) of replicated measurements performed on the same sample(s)—from one or preferably from several subjects—in the same run of analyses, in different runs or in different laboratories. The practice of using the coefficient of variation appears to be based on the fact that very often the measurement variability increases with the magnitude of the measurement, yielding a constant coefficient. Table 3 depicts the attenuation of a true relative risk (odds ratio) resulting from a method of measurement affected by errors expressed by three different coefficients of variation. For instance, with a coefficient of variation of 5%, a true odds ratio of 5.0 will be observed as a value of 4.56. Several assumptions have been made in constructing the table, the most relevant being that it refers to a variable (e.g. plasma total cholesterol) whose individual observed values in a population may exhibit a coefficient of variation of 15% around the population mean. For variables with a higher proportional dispersion the attenuation would be less, and for variables with a narrower proportional range of values in the population, the attenuation would be even more marked than it is in the table. Unless the coefficient of variation for the laboratory error is kept, say, within 5%, a material attenuation

Table 3. Observed (attenuated) odds ratios with an exposure measurement affected by laboratory random error

True odds ratio	Coefficient of variation of laboratory random error		
	3%	5%	10%
1.5	1.48	1.46	1.35
2	1.97	1.92	1.68
3	2.93	2.82	2.27
5	4.82	4.56	3.32
10	9.51	8.77	5.57

Table 4. Lung cancer occurrence in subjects classified by smoking status (S), orange juice intake (OJ) and vitamin C plasma levels

(a)

	OJ+ (n = 1000)	OJ– (n = 1000)
Lung cancer	60	60
No lung cancer	840	840

Relative risk (OJ+/OJ–) = 1.0

(b)

	S+ (n = 1000)		S– (n = 1000)	
	OJ+ (n = 500) Vitamin C low	OJ– (n = 500) Vitamin C very low	OJ+ (n = 500) Vitamin C high	OJ– (n = 500) Vitamin C low
Lung cancer	50	50	10	10
No lung cancer	450	450	490	490
	Relative risk (OJ+/OJ–) = 1.0		Relative risk (OJ+/OJ–)=1.0	

of the true odds ratio can ensue from this variation alone, to which further attenuation must be added, chiefly from intrasubject variability.

Metabolic confounding

A second aspect in need of thorough consideration concerns metabolic paths or, more generally, biological transformation processes (for instance, cell maturation processes) affecting a biomarker of exposure but not an external (questionnaire or environmental) measurement of exposure. Table 4(a) shows the lung cancer figures for 1000 consumers of orange juice (providing vitamin C supplementation) and 1000 non-consumers, the risk being the same in both categories. When stratified by smoking habit (Table 4b), the relative risk in the two categories of orange juice consumers remains 1; however, because of the enhancing effect of orange juice on vitamin C plasma levels (a biomarker), and the often reported lowering effect of smoking on the same biomarker, the measured levels of vitamin C are likely to turn out as schematized in the headings of Table 4(b) (high, low, very low). If one had categorized the biomarker levels without taking smoking habits into account, the exposure–risk relationship of Table 5 would have resulted, showing a clear increase of risk with decreasing levels of plasma vitamin C. Of course, tobacco smoke is such a well known source of metabolically active chemicals that information on it is very rarely omitted in epidemiological studies involving biomarkers, and correction for its confounding effect is feasible. Still, one may need better than only gross information on smoking habits; dose may matter, as may time, as some metabolic effects are long-term and others short-term. It may be less obvious that it is necessary to correct for the influences of other agents (drugs, nutrients), particularly as some metabolic chains, such as the enzymes of the P450 superfamily, may be interfered with by a vast number of xenobiotics, not all of which are known or measurable. In general, confounders of an exposure biomarker need not be the same as confounders of an alternative measurement of the same exposure. This is an important consideration, since, in the overall plan of what needs to be measured in a study, one has to include not only the exposure(s) of interest but key confounders as well.

Disease-induced differential error

This potential source of error is mentioned here for the sake of completeness, as it is usually—unlike other sources—well acknowledged in studies involving biomarkers. In case–control and cross-sectional investigations, the presence of disease,

Table 5. Lung cancer occurrence in subjects classified by Vitamin C plasma levels

	Vitamin C high (n = 500)	Vitamin C low (n = 1000)	Vitamin C very low (n = 500)
Lung cancer	10	60	50
No lung cancer	490	940	450
Relative risks	1.0 (Reference category)	3.1	5.4

e.g. colon cancer, may in fact be the cause of changes in a biomarker, e.g. blood cholesterol level, which will show up as an association between the biomarker and the disease. This association, due to differential error, could be wrongly interpreted as indicating that the biomarker precedes the disease in time and can therefore predict the disease occurrence. The same may apply in cohort studies, as an unrecognizable, still subclinical, disease may induce a change in a biomarker in those subjects who are bound, later on, to move on to clinically recognizable disease. Stratification of subjects by disease stages may be one way (often the only one) of getting a clue as to whether disease-induced differential error has occurred. An association which becomes stronger with advancing stages of the disease points to the possibility of a disease-induced differential error, although the absence of such a gradient represents only weak evidence against such bias, particularly if treatment has altered the natural course of the disease.

Sampling time variability

One well-recognized and critical aspect in the comparative evaluation of a biomarker with respect to an alternative measure of exposure is the time of appearance, persistence and disappearance of the biomarker in relation to the time of the external exposure or the fluctuations in time of an internal exposure (e.g. hormones). Time differences in sampling may contribute substantially to subject misclassification in respect to exposure, obscuring the detection of associated risks and reducing the superiority of the biomarker compared to alternative measurements. For example, if blood is taken from subjects during a survey lasting 3 months after a relevant exposure has occurred, for instance in an accident, and the half time of the biomarker of interest is of the same order of magnitude, a range of values of the marker from –30 to +30% of the average will be observed merely because of differences in the time of sampling, inducing a substantial misclassification of subjects by level of exposure. Alongside short-term (e.g. day-to-day or week-to-week) and medium-term (e.g. month-to-month) variations, longer term variations in a biomarker can occur over a period of years. For example, vitamin C intake could vary greatly over a person's life and it may be that intake over the 20 years before diagnosis influences lung cancer. In that case, a food frequency measure of vitamin C intake, which would have substantial error, could be a better measure of the relevant exposure than an accurate biomarker of vitamin C that measures only recent intake. Whereas the adverse effects of short-term and medium-term variation can be reduced by the use of multiple measurements, long-term variation in an exposure, and in the corresponding biomarker, seriously limits the use of the marker and can only be overcome if a marker of cumulative exposure becomes available.

In the case of internal exposure, hormones are known to exhibit sizeable daily variation (for instance, in the evening, levels of cortisol may be less than half that in the morning) and again casual sampling during the day may misclassify subjects and reduce the detectability of risks.

The general implication is that when assessing the total intrasubject and intersubject variability of a biomarker, a good deal can be gained by knowledge of its kinetics, which determines the magnitude of the differences related to sampling time.

Biomarkers of biological agents

In discussing the relative merits of biomarkers and alternative measurements of exposure, a basic truism

should not be forgotten: biomarkers are, by their very nature, at their best when measuring biological agents (exposures), provided the agent leaves—as is often the case because of the specific biological interactions with the host—a characteristic long-term trace in the body. For example, when, in a large cohort of more than 20 000 subjects, exposure to hepatitis B was assessed through anamnestic reporting of clinical hepatitis, a relative risk for subsequent primary hepatocellular carcinoma of about 4 was found; when, however, the HBAgs (Australia antigen) biomarker, indicative of a subject carrier status of the virus, was measured, the risk in subjects positive for the antigen was found to be more than 200 times higher than in subjects negative for the antigen, effectively ruling out confounding and making bias unlikely (Beasley *et al.*, 1981). Similarly, increased risks of cervical cancer associated with a history of sexual promiscuity, of the order of two- to threefold, were found in several studies (Cramer, 1982), leading to the hypothesis that a sexually transmissible biological agent was causally implied. Measuring exposure to papilloma viruses (particularly HPV 16) with increasingly reliable techniques, prevalence odds ratio of 20 and more have been shown in a large number of case–control studies, supporting the conclusion of a causal link between the virus and cervical cancer (IARC, 1995).

This advantage of biomarkers applies not only to the class of environmental biological agents but also to the other class of etiologically relevant biological agents, the genes. Indirectly assessing the exposure to a gene through family aggregation of cancer cases may cause an enormous attenuation of the true relative risk associated with a gene increasing the risk of cancer. Table 6 (from Peto, 1980) refers to the case of a gene with a dominant allele conferring an increased risk of cancer and having a population frequency of 5%. When those homo- and heterozygotes for the allele (95.7% of the population) are compared with those without the allele, a true relative risk of 100 becomes an observable relative risk of less than five if measured—as has been done for decades in epidemiology—by actually contrasting the cancer risk among siblings of cancer cases with the risk in the general population. This dilution, probably one of the strongest one can find in exposure assessment, derives from three sources: first, not all the cases of the investigated cancer have the risk-enhancing genotype; second, even fewer of their relatives have it; and third, the general population contains subjects both with and without that genotype. Now that molecular genetics techniques make it possible to determine directly a subject genotype, these sources of dilution have become virtually irrelevant, as comparisons of risks in persons with different genotypes, and different expressions of them, can be made. There is little question that genetic biomarkers, for use in both linkage and association studies, represent the most radical methodological advance for etiological studies in recent years.

Biomarkers as the reference

Comparisons between a biomarker and alternative measurements of exposure are made in three rather different circumstances:

Quantitative comparison

This comparison stands on the hypothesis that the two measurements measure the same quantity (exposure) and that, ideally, a correlation of 1.0 should exist between the two measurements. Divergence from this perfect correlation derives from errors in both measurements and can be quantified in the form of a reliability coefficient, which, besides being of interest in itself, allows the validity coefficient to be estimated, at least at its upper and lower bounds (Armstrong *et al.*, 1992, pp. 82–83). In the much less common situation in which sound evidence is available that the biomarker can be regarded as the ultimate and true

Table 6. Attenuation of the genetic relative risk of a cancer when investigated through the excess family risk

	Relative risk in relatives of cancer patients		
True relative risk (XX, OX versus OO)	Twin	Sibling	Child
2	1.07	1.04	1.04
10	3.02	2.00	1.99
100	8.60	4.75	4.70
1000	10.10	5.48	5.42

Table 7. Lung cancer occurrence in subjects classified by a marker composite of variables S (sensitivity = specificity = 100%) and BS (sensitivity = specificity = 80%)

	BS+, S+ (n = 680)	BS+, S–; BS–, S+ (n = 532)	BS–, S– (n = 788)
Lung cancer	64	22	8
No lung cancer	616	510	780
Relative risks	10.1	4.2	1.0 (Reference category)

measure of the exposure, both the validity and reliability of the alternative measure can be directly determined with respect to it.

Qualitative validation

This type of validation *assumes* the biomarker as the reference and validates the alternative measurement (obtained by questionnaire or environmental measurement) by looking for a correlation with the biomarker, the underlying and implicit hypothesis being that the correlation may indeed be zero, given the very indirect nature of the alternative measurement. This type of qualitative validation, in which finding a statistically significant correlation is even more important than the magnitude of the correlation, is particularly relevant when dealing with past exposures. For instance, that the levels of dioxin (TCDD) are higher in fat samples available from only a few pesticide sprayers than in the general population represents circumstantial evidence of the exposure to TCDD or to substances contaminated by TCDD. This, in turn, lends support to an interpretation in terms of occupational exposures of the relationship which may have been found between reconstructed work histories and disease end-points in the whole cohort of sprayers.

This recalls the story of the ancient city of Troy, in Asia Minor, as told in the epic poems of Homer. It was commonly held among scholars for generations that the events described by Homer, as they appeared so well detailed, never took place at all and were pure legend. However, in 1870, the German merchant and archaeologist Heinrich Schliemann went and dug where Homer said Troy was and found the site of the city (Lister, 1967). Incidentally, as he was using rather primitive archaeological techniques, he destroyed as many 'markers' as he revealed. From that moment, based on a limited validating evidence, the whole series of events and their interrelationship became regarded as the true history of the Troy war. In epidemiology, a 'Troy validation' of very indirect exposure measurements against the few feasible measurements of a meaningful biomarker may be formally unsatisfactory but substantively informative.

Analogic extrapolation

Whatever relationship is found between a biomarker and an alternative measure of exposure may be extrapolated, by analogy and with caution, to hold in circumstances different from those under which the relationship was found. For instance, in the short term (i.e. a week), reported exposure to environmental tobacco smoke in nonsmoking women correlates well with urinary cotinine levels, which reflects actual exposure to tobacco smoke (Riboli *et al.*, 1990). This offers support, by analogy, to the contention that in the long term (i.e. years), self-reported exposure to ETS reflects actual exposure to ETS.

Composite markers

One way in which biomarker measurements, even if imperfect, may be valuable is through combination into composite variables, formed by several biomarkers or by biomarkers and other measurements of exposure. This is particularly advantageous when some of the measurements can be performed cheaply. Referring back to the case of a random error affecting the measurement of a biomarker BS (Tables 1 and 2), assume now that the loss of sensitivity and specificity is sizeable, both being reduced to 80%. In these conditions, the relative

risk for the biomarker of tobacco smoke is just 3.4, while the relative risk assessed via a tobacco smoking questionnaire remains 7.4. Once the latter is known, the relative risk for the biomarker is, within the strata of tobacco smoking habits, only 1.8; it may be tempting (particularly if the biomarker measurement is costly) to dismiss it as worthless. This may in fact be wrong, as the biomarker measurement contributes information when considered jointly with the questionnaire measurement. An odds ratio of 10.1 is found in the highest category of a new variable created by compounding the questionnaire information and the biomarker information, which shows (Table 7) a clear gradient of risk with increasing levels of the variable. This elementary combination of variables underlines the important principle that when dealing with correlated variables purporting to measure the same underlying, but not directly accessible, variable, it may definitely be advantageous actually to measure several of them, even if they are individually affected by sizeable non-random errors, and to combine them in a composite variable (Armstrong *et al.*, 1992, pp. 115–125). Indeed, this may be one of the main ways of overcoming the inherent imperfection of exposure measurements in epidemiology.

Conclusion

Biomarkers of exposure are enriching the epidemiologist's armamentarium for evaluating exposures, present and past. The issue of the relative merits of biomarkers compared to alternative measures of exposure arises particularly in the context of studies aimed at hazard detection and quantification. In general, the relative merits of the different types of measurement of exposure need to be evaluated on a case-by-case basis. However, when the exposure is represented by a biological agent, biomarkers are the first, and often the only justifiable, choice. Whatever the type of study and the type of exposures involved, biomarkers may be affected by random errors, time-related sampling errors, physiological confounding and disease-induced differential error; full assessment of these aspects is needed before embarking on the use of a biomarker in a full-scale epidemiological study. Random errors affecting biomarkers may be reduced by replication or combination of measurements, or both. When evidence allows a biomarker to be regarded as the true measure of exposure, the validity and reliability of alternative measurements can be directly determined against it. A less elegant and ambitious, but no less valuable and common, use is in confirming, on the basis of whatever measurements may be feasible ('Troy validation'), that indeed other indirect measurements of exposure bear at least some relation to a biologically relevant exposure.

Two final remarks are apposite. First, there is no need to have biomarkers, as opposed to other exposure variables, measured in every epidemiological study, and epidemiologists ought to make clear and stress to colleagues from other disciplines, including those sitting in peer review and grant awarding committees, that what qualifies a good or a bad epidemiological study is not the mere presence or absence of measurements of some of the latest available biomarkers. Second, the flow of newly measurable biomarkers appears to be such that it is reasonable to expect that biomarkers of exposure which are capable of meeting the exacting requirements demanded by their use in epidemiology will soon be forthcoming.

References

Armstrong, B.K., White, E. & Saracci, R. (1992) *Principles of Exposure Measurement in Epidemiology.* Oxford, Oxford University Press

Beasley, R.P., Hwang, L.Y., Lin, C.C. & Chien, C.S. (1981) Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22707 men in Taiwan. *Lancet*, ii, 1129–1133

Cramer, D.W. (1982) Uterine cervix. In: Schottenfeld, D. & Fraumeni, J.F., eds, *Cancer Epidemiology and Prevention,* Philadelphia, W.B. Saunders, pp. 881–900

International Agency for Research on Cancer (1995) *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans,* Vol. 64, *Human Papillomaviruses,* Lyon, International Agency for Research on Cancer

Lister, R.P. (1967) *Turkey Observed,* London, Eyre & Spottiswoode

Peto, J. (1980) Genetic predisposition to cancer. In: Cairns, J. Lyon, J.L. & Skolnik, M., eds, *Cancer Incidence in Defined Populations,* Cold Spring Harbor, Cold Spring Harbor Laboratory, pp. 203–213

Phillips, A.N. & Davey Smith, G. (1991) How independent are "independent" effects? Relative risk estimation

when correlated exposures are measured imprecisely. *J. Clin. Epidemiol.*, 44, 1223–1231

Riboli, E., Preston-Martin, S., Saracci, R., Haley, N.J., Trichopoulos, D., Becher, H., Burch, J.D., Fontham, E.T.H., Gao, Y.T., Jindal, S.K., Koo, L.C., Le Marchand, L.L., Segnan, N., Shimizu, H., Stanta, G., Wu-Williams, A.H. & Zatonski W. (1990) Exposure of nonsmoking women to environmental tobacco smoke: a 10-country collaborative study. *Cancer Causes Control*, 1, 243–252

R. Saracci
National Research Council, Via Trieste, 41
56100 Pisa, Italy

Application of Biomarkers in Cancer Epidemiology
Toniolo, P., Boffetta, P., Shuker, D.E.G., Rothman, N., Hulka, B. and Pearce, N., eds
IARC Scientific Publications No. 142
International Agency for Research on Cancer, Lyon, 1997

Ethical and social issues in the use of biomarkers in epidemiological research

P.A. Schulte, D. Hunter and N. Rothman

The use of biomarkers in epidemiological research may raise ethical and social issues. These issues stem from the belief that research participants have 'rights' to appropriate information before, during and after studies so that they can make informed decisions. Ethical issues can arise during protocol development, obtaining participation, and in the interpretation and notification of text and study results. Additionally, there are ethical considerations concerning the use of biological specimens collected and stored for one purpose and subsequently used for other research purposes. A major ethical issue is the maintenance of participants' privacy and the confidentiality of their test and study results. Ethics committees need to be well-informed about the scope, limitations and expectations of biomarker research in order to be able to respond to social and scientific developments in the use of biomarkers.

Epidemiological research potentially raises many ethical questions and issues. The use of biomarkers in such research may raise further issues, because biomarkers are obtained from the individual person and have the potential for providing important information about exposures, biological effects of exposures and susceptibility to disease for that individual (Grandjean, 1991; Schulte, 1992; Van Damme *et al.*, 1995). At the same time, there is the widespread misconception that biological information is always more valid than other information, such as that obtained from questionnaires, environmental monitoring or record review. None the less, the potential contribution of biomarkers to enhancing determination of carcinogen exposure–disease associations, identifying disease earlier, or identifying particular etiological subgroups makes the use of biomarkers desirable and inevitable.

There is increasing recognition that many of the issues related to recruiting and informing subjects of test and study results have varied depending on study design. Consider three examples: (1) a cross-sectional study involving occupational exposure and a biomarker of early effect (e.g. cytogenetic effects); (2) a cancer case–control study evaluating the impact of common polymorphisms of metabolizing enzymes; and (3) a prospective cohort study with banked biological specimens. In general, cross-sectional studies of healthy workers are completed in a short period of time, with the expectation that the biomarkers under study may provide some insight into the potential risk of an exposed group as a whole, or possibly into an individual's risk of subsequent cancer development. Notifying workers of their results in these studies is common. By contrast, the case–control study involves subjects who are already sick, along with randomly selected controls who are not definable in an *a priori* sense to be at risk. These subjects are generally not notified of results. Finally, in the third example, everyone is healthy and samples are provided with the expectation that results will not be available for a relatively long time.

In the following pages, we will generalize about these issues, but the appropriateness may vary by study design or detail and the ethical issues should be addressed on a case-by-case basis.

In this paper, we will use the steps in the research process as the organizing theme and discuss ethical and social issues for each step. Where there are issues that differ according to the type of marker or the use of a marker, these aspects will be identified. Finally, we will discuss the use of stored specimens in biomarker research.

A premise of this paper is that ethical use of biomarkers in research involves attention to the 'rights' of subjects to appropriate information

before, during and after studies, so that they can make informed decisions. Failure to plan or budget adequately for these efforts can lead to these rights not being met. We would note, however, that there is some difference of opinion about when and what to tell research participants, and this will be discussed in later sections.

Protocol development and study design

Ethical issues come into play from the moment biomarkers are considered for a study. Why is the biomarker being considered? Biomarkers are usually more resource- and labour-intensive than other measures of exposure, outcome or risk. The use of scarce resources to develop, validate or apply a biomarker can be wasteful or inefficient if there is not a good rationale. Essential to the design of transitional, etiological or applied studies is the need to identify the driving scientific or public health questions and to determine whether they could be answered by some other approach (Rothman, 1993; Rothman *et al.*, 1995). This may be less of an issue for laboratory studies where biomarker work is the defining activity. It may be more critical when considering using biomarkers as independent or dependent variables in epidemiological studies or for public health applications such as screening, monitoring or in risk assessment (Office of Technology Assessment, 1990; Perera, 1987; Schulte & Halperin, 1987; Rüdiger, 1994).

Ethical and social problems may also arise from a failure of researchers to anticipate and plan the actions required for dealing with the more extreme biomarker assay results. This may include repeat testing, counselling or diagnostic evaluations. For transitional studies in which the characteristics of a marker are being determined, and where there are clearly no associated clinical findings, prognostic significance or meaning, the needs of subjects may be different from those situations, such as screening or biological monitoring, where a marker can have implications for individual risk or for disease. With markers of susceptibility, it may be important to consider the impact of the research not only on individual participants, but also on their families.

Obtaining participation

How subjects are recruited into studies can involve serious ethical and social issues (Schulte & Sweeney, 1995). These issues hinge on what potential subjects are told about the study and whether they can truly give informed consent. If subjects are deceived or coerced into participating in a study, or are given false expectations (e.g. 'we can tell if you are sick or well') with respect to the value of the study to the participant, ethical principles are violated. For example, a researcher could coerce a potential subject directly (e.g. 'you may lose your job if you don't participate') or by implication. Communicating false expectations or using pressure is patently dishonest and unethical. It is unlikely that such deception or coercion would be overt; rather, it would be more subtle and difficult to detect. A broad spectrum of opinion exists about what obtaining informed consent entails and when it is achieved. Some believe that for markers whose meaning is not known at the time of the study, a subject or worker in an occupational study cannot give truly informed consent (Samuels, 1994). This implies a much higher standard of interpretation for biomarker information than for other information routinely obtained by questionnaires, environment monitoring or record linkage. In studies to validate markers of exposure, the level of understanding of the meaning of the marker is similar to that from classical exposure sources. Frequently, airborne exposure, levels in blood, or frequency of DNA or protein adducts are part of the same exposure paradigm. Markers of effect or susceptibility are different. Until there is determination of predictive value and course in the natural history, such markers are clearly only research variables with no clinical meaning, and participants should be made aware of this. If a marker has been validated (i.e. quantitatively linked to risk of disease at the group or individual level), then a clear description of it should be given to potential research participants. With regard to informing participants of risks, general practice has been to identify only medical risks; however, it has been argued that truly informed consent should include reference to non-medical risks that might affect participants. For example, a study subject may be informed that they carry a genetic mutation that puts them at a high risk of subsequently developing cancer. In the extreme case, the mere acknowledgement on an employment or insurance application that they have had a biological or genetic test may result in denial of employment or insurance. Another variation on this scenario is that

misinterpretation of a biomarker assay result could occur and have the same impact.

Participants consent to provide the specimens and corollary demographic and risk factor information, and hence cooperate in the specified research. The subject generally does not consent or imply consent to distribution of the data in a way that identifies him or her individually to any other parties, such as employers, unions, insurers, credit agencies, lawyers, family members, public health agencies, etc.

Dissemination or revelation of results beyond the explicit purposes for which specimens were collected intrudes on subjects' privacy. Studies where biological specimens and DNA are banked for future use may require informed consent about this future use. In this respect, questions are raised about whether specimens collected for one purpose can be used for different research purposes and about the responsibility for conveying results back to the subjects (Schulte & Sweeney, 1995). Also related to this is the ownership of specimens. Who owns them—the subject, the researcher, the sponsoring agency or others? Although this has been adjudicated in the case of a clinician who profited from a hairy cell leukaemia line derived from cells taken from a patient (Cooper, 1985; Office of Technology Assessment, 1987), we have found no references (except Clayton *et al.*, 1995, see later discussion) to the issue as it pertains to epidemiological research with stored specimens.

Interpretation and notification of test and study results

Biomarker research yields individual test (assay) and study results (Schulte & Singal, 1989). Research participants may want, or have, a right to these results and an interpretation of them. Interpretation of these results is the responsibility of investigators. Some institutions require investigators to provide individual test results to subjects as well as overall study results, while others may advise them not to communicate results of assays that have no clinical relevance. Attendant to these efforts is the provision of an interpretation as far as is possible. Even though participants are told that tests may be purely for research purposes and have no clinical value, they still ultimately want to know if they are 'all right'. Investigators and practitioners face ethical issues in interpreting tests and in deciding when biomarkers indicate that early warning steps should be taken. These may include efforts to control exposures (in occupational or environmental settings), the need for subsequent testing, ongoing monitoring, or simply, and often most importantly, counselling and a demonstration of caring.

Interpretation of biomarker data can be difficult. For example, in cross-sectional studies of populations with occupational or environmental exposure, evaluating the relationship between exposure and markers of early biological effect, biomarkers will not be indicators of risk *per se*, but of exposure, susceptibility given exposure, or biological changes that could be homeostatic responses to an exposure (Ashford, 1986). The investigator needs to sort out these changes against a background of extensive intra-individual and interindividual variability in biomarkers. The current technological capabilities offer investigators and practitioners the opportunity to utilize techniques with heightened sensitivity for detecting changes at cellular and molecular levels, and for detecting exposures to minute amounts of a xenobiotic. At the same time, at these levels, inherited and acquired host factors and other confounding factors can be strong causes of wide variability in biomarker results unrelated to the exposure of interest.

The results of studies of biomarkers of susceptibility can lead to findings that might be misunderstood or abused (Lappe, 1983; Ashford, 1986; Nelkin & Tancredi, 1989). For example, some genes (such as those that are commonly occurring, that confer low relative risk and that require a specific exposure or other genes to increase risk of disease) (see Caporaso & Goldstein, this volume) do not provide unambiguous information, but various groups in society may start using such genotype information as if it represented 'diagnoses' rather than risk factors (Wagener, 1995).

In some studies, multiple biomarkers will be assessed, and researchers have a responsibility to consider whether issues of multiple comparisons can lead to inappropriate selection of significance levels. Association of biomarkers not included in original hypotheses should be evaluated at more rigorous levels of statistical significance, and subsequent interpretations should be considered in that light.

One area of interpretation that is problematic is what is called 'individual risk assessment'.

Generally, epidemiological studies (with or without biomarkers) yield group results. The risk pertains to the group as a whole and not necessarily to individual members of the group. It is possible to compute an individual risk using a risk function equation (Truett *et al.*, 1969); however, if the marker being used has not been validated for disease, the calculation will be meaningless. Thus far, for the current generation of molecular biomarkers used in cancer research, there are practically no markers, with the exception of a few genetic mutations linked to high risk of disease in cancer family syndromes, for which an individual risk can be determined based on the level of the marker.

All of these characteristics of biomarker data may lead an investigator to conclude that a particular biomarker is of uncertain meaning with regard to risk. None the less, the investigator has the obligation to portray accurately the degree of uncertainty in test and study results. There are a range of opinions about communicating results of biomarker tests on individuals or groups if there is no clinical meaning, such as usually occurs in transitional studies to validate markers. Some believe the autonomy of participants is not honoured if they do not receive the information, while others believe that the information is meaningless to participants. The latter view has the appearance of being paternalistic, but may be viewed as doing no harm.

Other ethical issues involved in notification are the importance of communicating information in a timely fashion and the evaluation of the impact of notification efforts. The timeliness of notification is mainly an issue when results indicate an action that could reduce exposure or risk, or effect timely treatment. Evaluating the impact of notifications may not need to be a routine matter, but since the impact of notification cannot always be anticipated, it may be useful to have included in the notification an opportunity for the participant to obtain more information or provide feedback about the results.

Use of stored specimens in biomarker research

Biomarker research is qualitatively different from most other epidemiological research, because technical developments make new assays feasible on stored specimens long after the original consent is obtained. Unlike questionnaire-based research, in which the response to a new hypothesis is usually to start a new study and ask the relevant questions, a new hypothesis using a biomarker can often be tested using specimens from previous studies. If it is desirable to have prospectively collected specimens, for instance if the biomarker level may be biased by disease, then available specimen banks with follow-up data will be the preferred resource for testing the new hypothesis. Otherwise, it might take many years to develop a new specimen bank with sufficient outcomes and follow-up to test a hypothesis.

Ethical issues for stored specimens relate to (1) whether consent for use of the specimens in research was originally given, and (2) whether this consent was generic or specific to the hypothesis to be tested, and whether the consent obtained when the specimens were collected still meets the standards of informed consent.

Many specimens stored for research purposes would have been collected after informed consent to research was given; however, some types of specimens, particularly clinical specimens initially used for diagnostic or prognostic purposes, may have been stored without consent or even without the patient's knowledge. Frequently in clinical settings, a wide variety of tests are ordered without any consultation with the patient, although clear exceptions exist, such as HIV testing, for which consent is usually mandatory. It has long been held as ethically acceptable practice to conduct some types of research on 'discarded' blood or tissues, i.e. specimens left over after the clinical tests are performed. Access to these tissues has been critical to the development of new clinical markers such as histological or immunochemical markers of cancer prognosis, in which hundreds or thousands of uniformly collected specimens are frequently needed to establish a new test as being informative. It would seem a natural extension of this tradition that new biomarkers of genetic susceptibility or prognosis would also be evaluated in this way. However, because of the potential high predictive value of some of these tests, as well as the implications for family members, this tradition is being challenged, and a lively debate is currently underway about the ethics of using these tissues. A recent statement from a working group of the Ethical, Legal, and Social Implications of the Human Genome Project suggested that informed consent should usually be obtained before testing for genetic susceptibility on clinical specimens

(Clayton *et al.*, 1995), although the statement acknowledged that research involving 'minimal risk', and for which re-consenting subjects would be impracticable, could be exempted. The definition of 'minimal risk' and the determination of what constitutes 'impracticability' are at the centre of much of the current uncertainty and debate.

Even in a research study, the original consent form can only be as thorough as the original aims of the study and the state of knowledge at the time permit. Samples from participants in a study of cancer risk factors, for instance, may subsequently be useful in a study of cardiovascular disease or psychiatric illness. Even the best designed and informed consent process in a study of genetic susceptibility to cancer may be outdated with the discovery of a new susceptibility gene or a new prognostic implication of an 'old' gene. A major dilemma in current biomarker research is whether the generic consent originally given by a participant to do research on a specimen is adequate consent to conduct a specific test which may not even have been envisaged at the start of the study. The obvious strategy of obtaining fresh consent has at least three major problems: (1) subjects may be very difficult to contact if follow-up has not been maintained, or they may have died; (2) a high proportion of non-consent, due either to inability to re-contact or to refusal, may bias the study; (3) for certain especially valuable specimens, such as those from cohort studies, multiple genes may be of interest and a process of very specific informed consent would generate an almost continuous stream of consent requests to the participant. Failure to obtain a new informed consent may expose the researcher to allegations of unethical behaviour, or may create a difficult situation if the biomarker information is of clinical relevance to the participant and yet the participant was not pre-test counselled about the test. The nature and force of these problems will be very different according to the predictive quality of the biomarker and its clinical implications, and the social and cultural setting of the research.

Owing to the heterogeneity of study settings, and of social norms and responses, it is likely to be impossible to draft uniform rules on what constitutes ethical behaviour in every application of biomarker research and every situation. This is currently the case with research involving human subjects, in which the first rule is that virtually all such research must be approved and reviewed by an appropriate ethics committee but relatively few types of research are absolutely proscribed or highly regulated. Some have proposed that research involving genetic susceptibility is qualitatively different from other research, and that much stricter standards of informed consent should apply (Annas, 1995); while others have argued that the level of consent or notification should be commensurate with the degree of risk involved, and thus less stringent procedures may be appropriate for low risk, relatively common polymorphisms (e.g. *P450* genes) than for high-risk genotypes (e.g. *BRCA1*, *BRCA2*). In the USA, the possibility that biomarkers of susceptibility could be used to discriminate in the context of health insurance or employment is a major concern, which may expose research participants to potential economic harm. On the other hand, epidemiology has a good track record in protecting participants from loss of confidentiality in many studies over many years which have included highly sensitive questionnaire-based data. Although some unique issues are raised by biomarker research, most issues are similar to those encountered in other types of research, and can be overseen by appropriately constituted ethics committees who are in the best position to be aware of the local and particular aspects of any proposed biomarker research. Especially close scrutiny should be given to any proposal using a biomarker with likely high predictive value. Ethics committees also need to be well informed about the scope, limitations and implications of biomarker research, as the ethical climate in this field may change quite rapidly as scientific developments occur and society responds to these developments.

Confidentiality of data

Investigators need to maintain the confidentiality of biomarker data because of the potential for misuse or abuse leading to discrimination, labelling and stigmatization. This can be increasingly difficult because ownership of stored specimens may be in question and various investigators may request the use of specimens for research, litigation or commercial enterprise. In some cases, where specimens are identifiable or are capable of being linked to databases where identification is possible, it may be difficult to assure confidentiality.

Informatics and the ability to link disparate databases are progressing at a rapid pace. In some countries, there may be a need for further legislation to prohibit unauthorized access to, or use of, specimen results. The challenge will be to assure the rights of study participants while providing for a broad range of research opportunities.

References

Annas, G.J. (1995) Editorial: Genetic prophecy and genetic privacy—can we prevent the dream from becoming a nightmare? *Am. J. Public Health*, 85, 1196–1197

Ashford, N.A. (1986) Medical screening in the workplace: legal and ethical considerations. *Semin. Occup. Med.*, 1, 67–79

Clayton, E.W., Steinberg, K.K., Khoury, M.J., Thomson, E., Andrews, L., Kahn , M.E., Kopelman, L.M. & Weiss, J.O. (1995) Informed consent for genetic research on stored tissue samples. *J. Am. Med. Assoc.*, 274, 1786–1792

Cooper, J.P. (1985). *Biotechnology and the Law*, New York, Clark Boardman Ltd

Grandjean, P. (1991) Ethical aspects of genetic predisposition to disease. In: Grandjean, P., ed, *Ecogenetics*, London, Chapman and Hall, pp. 237–251

Lappe, M. (1983) Ethical issues in testing for differential sensitivity to occupational hazards. *J. Occup. Med.*, 25, 797–808

Nelkin, D. & Tancredi, L. (1989) *Dangerous Diagnostics. The Social Power of Biological Information*, New York, Basic Books, Inc

Office of Technology Assessment (1987) *New Developments in Biotechnology*, Vol. 1, *Ownership of Human Tissues and Cells* (OTA-BA-337), Washington, US Government Printing Office

Perera, F.P. (1987) The potential usefulness of biological markers in risk assessment. *Environ. Health Perspect.*, 76, 141–145

Rothman, N. (1993) Epilogue. In: Schulte, P.A. & Perera, F.V., eds, *Molecular Epidemiology: Principles and Practices*, San Diego, Academic Press, pp. 565–569

Rothman N. Stewart, W.F. & Schulte, P.A. (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol. Biomarkers Prev.*, 4, 301–311

Rüdiger, H.W. (1994) Changing emphasis in occupational toxicology: are we ready? *Int. Arch. Occup. Env. Health*, 66, 69

Samuels, S.W. (1994) A moral history of the evolution of a caste of workers. Presentation at Workshop on the Development and Applications of Biomarkers, Santa Fe, NM, April 26–29

Schulte, P.A. (1992) Biomarkers in epidemiology: Scientific issues and ethical implications. *Environ. Health Perspect.*, 98, 143–147

Schulte, P.A. & Halperin, W.E. (1987) Genetic screening and monitoring for workers. In: Harrington, J.M., ed., *Recent Advances in Occupational Health*, Vol. 3, Edinburgh, Churchill Livingstone, pp. 135–154

Schulte, P.A. & Singal, M. (1989) Interpretation and communication of the results of medical field investigations. *J. Occup. Med.*, 31, 5898–5894

Schulte, P.A. & Sweeney, M.H. (1995) Ethical considerations, confidentiality issues, rights of human subjects, and uses of monitoring data in research and regulation. *Environ. Health Perspect.*, 103 (suppl. 3), 69–74

Truett, J., Cornfield, J. & Kannel, W. (1967) A multivariate analysis of coronary heart disease in Framingham. *J. Chron. Dis.*, 20, 511–524

Van Damme, K., Casteleyn, L., Heseltine, E., Huici, A., Sorsa, M., Larebeko, N. & Vineis, P. (1995) Individual susceptibility and prevention of occupational diseases: scientific and ethical issues. *J. Occup. Environ. Med.*, 37, 91–99

Wagener, D.K. (1995) Ethical considerations in the design and execution of the national and Hispanic Health and Nutrition Examination Survey (HANES). *Environ. Health Perspect.*, 103 (suppl. 3), 75–80

Corresponding author

P.A. Schulte

National Institute for Occupational Safety and Health, 4676 Columbia Parkway, C-14, Cincinnati, OH 45226, USA

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans

Volume 1
Some Inorganic Substances, Chlorinated Hydrocarbons, Aromatic Amines, N-Nitroso Compounds, and Natural Products
1972; 184 pages; ISBN 92 832 1201 0
(out of print)

Volume 2
Some Inorganic and Organometallic Compounds
1973; 181 pages; ISBN 92 832 1202 9
(out of print)

Volume 3
Certain Polycyclic Aromatic Hydrocarbons and Heterocyclic Compounds
1973; 271 pages; ISBN 92 832 1203 7
(out of print)

Volume 4
Some Aromatic Amines, Hydrazine and Related Substances, N-Nitroso Compounds and Miscellaneous Alkylating Agents
1974; 286 pages; ISBN 92 832 1204 5

Volume 5
Some Organochlorine Pesticides
1974; 241 pages; ISBN 92 832 1205 3
(out of print)

Volume 6
Sex Hormones
1974; 243 pages; ISBN 92 832 1206 1
(out of print)

Volume 7
Some Anti-Thyroid and Related Substances, Nitrofurans and Industrial Chemicals
1974; 326 pages; ISBN 92 832 1207 X
(out of print)

Volume 8
Some Aromatic Azo Compounds
1975; 357 pages; ISBN 92 832 1208 8

Volume 9
Some Aziridines, N-, S- and O-Mustards and Selenium
1975; 268 pages; ISBN 92 832 1209 6

Volume 10
Some Naturally Occurring Substances
1976; 353 pages; ISBN 92 832 1210 X
(out of print)

Volume 11
Cadmium, Nickel, Some Epoxides, Miscellaneous Industrial Chemicals and General Considerations on Volatile Anaesthetics
1976; 306 pages; ISBN 92 832 1211 8
(out of print)

Volume 12
Some Carbamates, Thiocarbamates and Carbazides
1976; 282 pages; ISBN 92 832 1212 6

Volume 13
Some Miscellaneous Pharmaceutical Substances
1977; 255 pages; ISBN 92 832 1213 4

Volume 14
Asbestos
1977; 106 pages; ISBN 92 832 1214 2
(out of print)

Volume 15
Some Fumigants, the Herbicides 2,4-D and 2,4,5-T, Chlorinated Dibenzodioxins and Miscellaneous Industrial Chemicals
1977; 354 pages; ISBN 92 832 1215 0
(out of print)

Volume 16
Some Aromatic Amines and Related Nitro Compounds – Hair Dyes, Colouring Agents and Miscellaneous Industrial Chemicals
1978; 400 pages; ISBN 92 832 1216 9

Volume 17
Some N-Nitroso Compounds
1978; 365 pages; ISBN 92 832 1217 7

Volume 18
Polychlorinated Biphenyls and Polybrominated Biphenyls
1978; 140 pages; ISBN 92 832 1218 5

Volume 19
Some Monomers, Plastics and Synthetic Elastomers, and Acrolein
1979; 513 pages; ISBN 92 832 1219 3
(out of print)

Volume 20
Some Halogenated Hydrocarbons
1979; 609 pages; ISBN 92 832 1220 7
(out of print)

Volume 21
Sex Hormones (II)
1979; 583 pages; ISBN 92 832 1521 4

Volume 22
Some Non-Nutritive Sweetening Agents
1980; 208 pages; ISBN 92 832 1522 2

Volume 23
Some Metals and Metallic Compounds
1980; 438 pages; ISBN 92 832 1523 0
(out of print)

Volume 24
Some Pharmaceutical Drugs
1980; 337 pages; ISBN 92 832 1524 9

Volume 25
Wood, Leather and Some Associated Industries
1981; 412 pages; ISBN 92 832 1525 7

Volume 26
Some Antineoplastic and Immunosuppressive Agents
1981; 411 pages; ISBN 92 832 1526 5

Volume 27
Some Aromatic Amines, Anthraquinones and Nitroso Compounds, and Inorganic Fluorides Used in Drinking Water and Dental Preparations
1982; 341 pages; ISBN 92 832 1527 3

Volume 28
The Rubber Industry
1982; 486 pages; ISBN 92 832 1528 1

Volume 29
Some Industrial Chemicals and Dyestuffs
1982; 416 pages; ISBN 92 832 1529 X

Volume 30
Miscellaneous Pesticides
1983; 424 pages; ISBN 92 832 1530 3

Volume 31
Some Food Additives, Feed Additives and Naturally Occurring Substances
1983; 314 pages; ISBN 92 832 1531 1

Volume 32
Polynuclear Aromatic Compounds, Part 1: Chemical, Environmental and Experimental Data
1983; 477 pages; ISBN 92 832 1532 X

Volume 33
Polynuclear Aromatic Compounds, Part 2: Carbon Blacks, Mineral Oils and Some Nitroarenes
1984; 245 pages; ISBN 92 832 1533 8
(out of print)

Volume 34
Polynuclear Aromatic Compounds, Part 3: Industrial Exposures in Aluminium Production, Coal Gasification, Coke Production, and Iron and Steel Founding
1984; 219 pages; ISBN 92 832 1534 6

Volume 35
Polynuclear Aromatic Compounds: Part 4: Bitumens, Coal-Tars and Derived Products, Shale-Oils and Soots
1985; 271 pages; ISBN 92 832 1535 4

Volume 36
Allyl Compounds, Aldehydes, Epoxides and Peroxides
1985; 369 pages; ISBN 92 832 1536 2

Volume 37
Tobacco Habits Other than Smoking; Betel-Quid and Areca-Nut Chewing; and Some Related Nitrosamines
1985; 291 pages; ISBN 92 832 1537 0

Volume 38
Tobacco Smoking
1986; 421 pages; ISBN 92 832 1538 9

Volume 39
Some Chemicals Used in Plastics and Elastomers
1986; 403 pages; ISBN 92 832 1239 8

Volume 40
Some Naturally Occurring and Synthetic Food Components, Furocoumarins and Ultraviolet Radiation
1986; 444 pages; ISBN 92 832 1240 1

Volume 41
Some Halogenated Hydrocarbons and Pesticide Exposures
1986; 434 pages; ISBN 92 832 1241 X

Volume 42
Silica and Some Silicates
1987; 289 pages; ISBN 92 832 1242 8

Volume 43
Man-Made Mineral Fibres and Radon
1988; 300 pages; ISBN 92 832 1243 6

Volume 44
Alcohol Drinking
1988; 416 pages; ISBN 92 832 1244 4

Volume 45
Occupational Exposures in Petroleum Refining; Crude Oil and Major Petroleum Fuels
1989; 322 pages; ISBN 92 832 1245 2

Volume 46
Diesel and Gasoline Engine Exhausts and Some Nitroarenes
1989; 458 pages; ISBN 92 832 1246 0

Volume 47
Some Organic Solvents, Resin Monomers and Related Compounds, Pigments and Occupational Exposures in Paint Manufacture and Painting
1989; 535 pages; ISBN 92 832 1247 9

Volume 48
Some Flame Retardants and Textile Chemicals, and Exposures in the Textile Manufacturing Industry
1990; 345 pages; ISBN: 92 832 1248 7

Volume 49
Chromium, Nickel and Welding
1990; 677 pages; ISBN: 92 832 1249 5

Volume 50
Some Pharmaceutical Drugs
1990; 415 pages; ISBN: 92 832 1259 9

Volume 51
Coffee, Tea, Mate, Methylxanthines and Methylglyoxal
1991; 513 pages; ISBN: 92 832 1251 7

Volume 52
Chlorinated Drinking-Water; Chlorination By-products; Some other Halogenated Compounds; Cobalt and Cobalt Compounds
1991; 544 pages; ISBN: 92 832 1252 5

Volume 53
Occupational Exposures in Insecticide Application, and Some Pesticides
1991; 612 pages; ISBN 92 832 1253 3

Volume 54
Occupational Exposures to Mists and Vapours from Strong Inorganic Acids; and other Industrial Chemicals
1992; 336 pages; ISBN 92 832 1254 1

Volume 55
Solar and Ultraviolet Radiation
1992; 316 pages; ISBN 92 832 1255 X

Volume 56
Some Naturally Occurring Substances: Food Items and Constituents, Heterocyclic Aromatic Amines and Mycotoxins
1993; 600 pages; ISBN 92 832 1256 8

Volume 57
Occupational Exposures of Hairdressers and Barbers and Personal Use of Hair Colourants; Some Hair Dyes, Cosmetic Colourants, Industrial Dyestuffs and Aromatic Amines
1993; 428 pages; ISBN 92 832 1257 6

Volume 58
Beryllium, Cadmium, Mercury and Exposures in the Glass Manufacturing Industry
1994; 444 pages; ISBN 92 832 1258 4

Volume 59
Hepatitis Viruses
1994; 286 pages; ISBN 92 832 1259 2

Volume 60
Some Industrial Chemicals
1994; 560 pages; ISBN 92 832 1260 6

Volume 61
Schistosomes, Liver Flukes and *Helicobacter pylori*
1994; 280 pages; ISBN 92 832 1261 4

Volume 62
Wood Dusts and Formaldehyde
1995; 405 pages; ISBN 92 832 1262 2

Volume 63
Dry cleaning, Some Chlorinated Solvents and Other Industrial Chemicals
1995; 558 pages; ISBN 92 832 1263 0

Volume 64
Human Papillomaviruses
1995; 409 pages; ISBN 92 832 1264 9

Volume 65
Printing Processes, Printing Inks, Carbon Blacks and Some Nitro Compounds
1996; 578 pages; ISBN 92 832 1265 7

Volume 66
Some Pharmaceutical Drug
1996; 514 pages; ISBN 92 832 1266 5

Volume 67
Human Immunodeficiency Viruses and Human T-cell Lymphotropic Viruses
1996; 424 pages; ISBN 92 832 1267 3

Volume 68
Silica, Some Silicates, Coal Dust and para-Aramid Fibrils
1997; 506 pages; ISBN 92 832 1268 1

Volume 69
Polychlorinated Dibenzo-para-dioxins and Dibenzofurans
1997; c. 600 pages; ISBN 92 832 1269 X

Supplements

Supplement No.1
Chemicals and Industrial Processes Associated with Cancer in Humans (IARC Monographs, Volumes 1 to 20)
1979; 71 pages; ISBN 92 832 1404 8
(out of print)

Supplement No. 2
Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal
1980; 426 pages; ISBN 92 832 1404 8

Supplement No. 3
Cross Index of Synonyms and Trade Names in Volumes 1 to 26
1982; 199 pages; ISBN 92 832 1405 6
(out of print)

Supplement No.4
Chemicals, Industrial Processes and Industries Associated with Cancer in Humans (Volumes 1 to 29)
1982; 292 pages; ISBN 92 832 1407 2
(out of print)

Supplement No. 5
Cross Index of Synonyms and Trade Names in Volumes 1 to 36
1985; 259 pages; ISBN 92 832 1408 0
(out of print)

Supplement No. 6
Genetic and Related Effects: An Updating of Selected IARC Monographs from Volumes 1 to 42
1987; 729 pages; ISBN 92 832 1409 9

Supplement No. 7
Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1 to 42
1987; 440 pages; ISBN 92 832 1411 0

Supplement No. 8
Cross Index of Synonyms and Trade Names in Volumes 1 to 46
1989; 346 pages; ISBN 92 832 1417 X

IARC Scientific Publications

No. 1
Liver Cancer
1971; 176 pages; ISBN 0 19 723000 8

No. 2
Oncogenesis and Herpesviruses
Edited by P.M. Biggs, G. de Thé and L.N. Payne
1972; 515 pages; ISBN 0 19 723001 6

No. 3
N-Nitroso Compounds: Analysis and Formation
Edited by P. Bogovski, R. Preussman and E.A. Walker
1972; 140 pages; ISBN 0 19 723002 4

No. 4
Transplacental Carcinogenesis
Edited by L. Tomatis and U. Mohr
1973; 181 pages; ISBN 0 19 723003 2

No. 5/6
Pathology of Tumours in Laboratory Animals. Volume 1: Tumours of the Rat
Edited by V.S. Turusov
1973/1976; 533 pages; ISBN 92 832 1410 2

No. 7
Host Environment Interactions in the Etiology of Cancer in Man
Edited by R. Doll and I. Vodopija
1973; 464 pages; ISBN 0 19 723006 7

No. 8
Biological Effects of Asbestos
Edited by P. Bogovski, J.C. Gilson, V. Timbrell and J.C. Wagner
1973; 346 pages; ISBN 0 19 723007 5

No. 9
N-Nitroso Compounds in the Environment
Edited by P. Bogovski and E.A. Walker
1974; 243 pages; ISBN 0 19 723008 3

No. 10
Chemical Carcinogenesis Essays
Edited by R. Montesano and L. Tomatis
1974; 230 pages; ISBN 0 19 723009 1

No. 11
Oncogenesis and Herpes-viruses II
Edited by G. de-Thé, M.A. Epstein and H. zur Hausen
1975; Two volumes, 511 pages and 403 pages; ISBN 0 19 723010 5

No. 12
Screening Tests in Chemical Carcinogenesis
Edited by R. Montesano, H. Bartsch and L. Tomatis
1976; 666 pages; ISBN 0 19 723051 2

No. 13
Environmental Pollution and Carcinogenic Risks
Edited by C. Rosenfeld and W. Davis
1975; 441 pages; ISBN 0 19 723012 1

No. 14
Environmental N-Nitroso Compounds. Analysis and Formation
Edited by E.A. Walker, P. Bogovski and L. Griciute
1976; 512 pages; ISBN 0 19 723013 X

No. 15
Cancer Incidence in Five Continents, Volume III
Edited by J.A.H. Waterhouse, C. Muir, P. Correa and J. Powell
1976; 584 pages; ISBN 0 19 723014 8

No. 16
Air Pollution and Cancer in Man
Edited by U. Mohr, D. Schmähl and L. Tomatis
1977; 328 pages; ISBN 0 19 723015 6

No. 17
Directory of On-Going Research in Cancer Epidemiology 1977
Edited by C.S. Muir and G. Wagner
1977; 599 pages; ISBN 92 832 1117 0
(out of print)

No. 18
Environmental Carcinogens. Selected Methods of Analysis. Volume 1: Analysis of Volatile Nitrosamines in Food
Editor-in-Chief: H. Egan
1978; 212 pages; ISBN 0 19 723017 2

No. 19
Environmental Aspects of N-Nitroso Compounds
Edited by E.A. Walker, M. Castegnaro, L. Griciute and R.E. Lyle
1978; 561 pages; ISBN 0 19 723018 0

No. 20
Nasopharyngeal Carcinoma: Etiology and Control
Edited by G. de Thé and Y. Ito
1978; 606 pages; ISBN 0 19 723019 9

No. 21
Cancer Registration and its Techniques
Edited by R. MacLennan, C. Muir, R. Steinitz and A. Winkler
1978; 235 pages; ISBN 0 19 723020 2

No. 22
Environmental Carcinogens: Selected Methods of Analysis. Volume 2: Methods for the Measurement of Vinyl Chloride in Poly(vinyl chloride), Air, Water and Foodstuffs
Editor-in-Chief: H. Egan
1978; 142 pages; ISBN 0 19 723021 0

No. 23
Pathology of Tumours in Laboratory Animals. Volume II: Tumours of the Mouse
Editor-in-Chief: V.S. Turusov
1979; 669 pages; ISBN 0 19 723022 9

No. 24
Oncogenesis and Herpesviruses III
Edited by G. de-Thé, W. Henle and F. Rapp
1978; Part I: 580 pages, Part II: 512 pages; ISBN 0 19 723023 7

No. 25
Carcinogenic Risk: Strategies for Intervention
Edited by W. Davis and C. Rosenfeld
1979; 280 pages; ISBN 0 19 723025 3

No. 26
Directory of On-going Research in Cancer Epidemiology 1978
Edited by C.S. Muir and G. Wagner
1978; 550 pages; ISBN 0 19 723026 1
(out of print)

No. 27
Molecular and Cellular Aspects of Carcinogen Screening Tests
Edited by R. Montesano, H. Bartsch and L. Tomatis
1980; 372 pages; ISBN 0 19 723027 X

No. 28
Directory of On-going Research in Cancer Epidemiology 1979
Edited by C.S. Muir and G. Wagner
1979; 672 pages; ISBN 92 832 1128 6
(out of print)

No. 29
Environmental Carcinogens. Selected Methods of Analysis. Volume 3: Analysis of Polycyclic Aromatic Hydrocarbons in Environmental Samples
Editor-in-Chief: H. Egan
1979; 240 pages; ISBN 0 19 723028 8

No. 30
Biological Effects of Mineral Fibres
Editor-in-Chief: J.C. Wagner
1980; Two volumes, 494 pages & 513 pages; ISBN 0 19 723030 X

No. 31
N-Nitroso Compounds: Analysis, Formation and Occurrence
Edited by E.A. Walker, L. Griciute, M. Castegnaro and M. Börzsönyi
1980; 835 pages; ISBN 0 19 723031 8

No. 32
Statistical Methods in Cancer Research.Volume 1: The Analysis of Case-control Studies
By N.E. Breslow and N.E. Day
1980; 338 pages; ISBN 92 832 0132 9

No. 33
Handling Chemical Carcinogens in the Laboratory
Edited by R. Montesano, H. Bartsch, E. Boyland, G. Della Porta, L. Fishbein, R.A. Griesemer, A.B. Swan and L. Tomatis
1979; 32 pages; ISBN 0 19 723033 4
(out of print)

No. 34
Pathology of Tumours in Laboratory Animals. Volume III: Tumours of the Hamster
Editor-in-Chief: V.S. Turusov
1982; 461 pages; ISBN 0 19 723034 2

No. 35
Directory of On-going Research in Cancer Epidemiology 1980
Edited by C.S. Muir and G. Wagner
1980; 660 pages; ISBN 0 19 723035 0
(out of print)

No. 36
Cancer Mortality by Occupation and Social Class 1851–1971
Edited by W.P.D. Logan
1982; 253 pages; ISBN 0 19 723036 9

No. 37
Laboratory Decontamination and Destruction of Aflatoxins B1, B2, G1, G2 in Laboratory Wastes
Edited by M. Castegnaro, D.C. Hunt, E.B. Sansone, P.L. Schuller, M.G. Sirlwardana, G.M. Telling, H.P. van Egmond and E.A. Walker
1980; 56 pages; ISBN 0 19 723037 7

No. 38
Directory of On-going Research in Cancer Epidemiology 1981
Edited by C.S. Muir and G. Wagner
1981; 696 pages; ISBN 0 19 723038 5
(out of print)

No. 39
Host Factors in Human Carcinogenesis
Edited by H. Bartsch and B. Armstrong
1982; 583 pages;
ISBN 0 19 723039 3

No. 40
Environmental Carcinogens: Selected Methods of Analysis. Volume 4: Some Aromatic Amines and Azo Dyes in the General and Industrial Environment
Edited by L. Fishbein, M. Castegnaro, I.K. O'Neill and H. Bartsch
1981; 347 pages; ISBN 0 19 723040 7

No. 41
N-Nitroso Compounds: Occurrence and Biological Effects
Edited by H. Bartsch, I.K. O'Neill, M. Castegnaro and M. Okada
982; 755 pages; ISBN 0 19 723041 5

No. 42
Cancer Incidence in Five Continents Volume IV
Edited by J. Waterhouse, C. Muir, K. Shanmugaratnam and J. Powell
1982; 811 pages; ISBN 0 19 723042 3

No. 43
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some N-Nitrosamines
Edited by M. Castegnaro, G. Eisenbrand, G. Ellen, L. Keefer, D. Klein, E.B. Sansone, D. Spincer, G. Telling and K. Webb
1982; 73 pages; ISBN 0 19 723043 1

No. 44
Environmental Carcinogens: Selected Methods of Analysis.
Volume 5: Some Mycotoxins
Edited by L. Stoloff, M. Castegnaro, P. Scott, I.K. O'Neill and H. Bartsch
1983; 455 pages; ISBN 0 19 723044 X

No. 45
Environmental Carcinogens: Selected Methods of Analysis.
Volume 6: N-Nitroso Compounds
Edited by R. Preussmann, I.K. O'Neill, G. Eisenbrand, B. Spiegelhalder and H. Bartsch
1983; 508 pages; ISBN 0 19 723045 8

No. 46
Directory of On-going Research in Cancer Epidemiology 1982
Edited by C.S. Muir and G. Wagner
1982; 722 pages; ISBN 0 19 723046 6
(out of print)

No. 47
Cancer Incidence in Singapore 1968–1977
Edited by K. Shanmugaratnam, H.P. Lee and N.E. Day
1983; 171 pages; ISBN 0 19 723047 4

No. 48
Cancer Incidence in the USSR (2nd Revised Edition)
Edited by N.P. Napalkov, G.F. Tserkovny, V.M. Merabishvili, D.M. Parkin, M. Smans and C.S. Muir
1983; 75 pages; ISBN 0 19 723048 2

No. 49
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Polycyclic Aromatic Hydrocarbons
Edited by M. Castegnaro, G. Grimmer, O. Hutzinger, W. Karcher, H. Kunte, M. Lafontaine, H.C. Van der Plas, E.B. Sansone and S.P. Tucker
1983; 87 pages; ISBN 0 19 723049 0

No. 50
Directory of On-going Research in Cancer Epidemiology 1983
Edited by C.S. Muir and G. Wagner
1983; 731 pages; ISBN 0 19 723050 4
(out of print)

No. 51
Modulators of Experimental Carcinogenesis
Edited by V. Turusov and R. Montesano
1983; 307 pages; ISBN 0 19 723060 1

No. 52
Second Cancers in Relation to Radiation Treatment for Cervical Cancer: Results of a Cancer Registry Collaboration
Edited by N.E. Day and J.C. Boice, Jr
1984; 207 pages; ISBN 0 19 723052 0

No. 53
Nickel in the Human Environment
Editor-in-Chief: F.W. Sunderman, Jr
1984; 529 pages; ISBN 0 19 723059 8

No. 54
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Hydrazines
Edited by M. Castegnaro, G. Ellen, M. Lafontaine, H.C. van der Plas, E.B. Sansone and S.P. Tucker
1983; 87 pages; ISBN 0 19 723053

No. 55
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some N-Nitrosamides
Edited by M. Castegnaro, M. Bernard, L.W. van Broekhoven, D. Fine, R. Massey, E.B. Sansone, P.L.R. Smith, B. Spiegelhalder, A. Stacchini, G. Telling and J.J. Vallon
1984; 66 pages; ISBN 0 19 723054 7

No. 56
Models, Mechanisms and Etiology of Tumour Promotion
Edited by M. Börzsönyi, N.E. Day, K. Lapis and H. Yamasaki
1984; 532 pages; ISBN 0 19 723058 X

No. 57
N-Nitroso Compounds: Occurrence, Biological Effects and Relevance to Human Cancer
Edited by I.K. O'Neill, R.C. von Borstel, C.T. Miller, J. Long and H. Bartsch
1984; 1013 pages; ISBN 0 19 723055 5

No 58
Age-related Factors in Carcinogenesis
Edited by A. Likhachev, V. Anisimov and R. Montesano
1985; 288 pages; ISBN 92 832 1158 8

No. 59
Monitoring Human Exposure to Carcinogenic and Mutagenic Agents
Edited by A. Berlin, M. Draper, K. Hemminki and H. Vainio
1984; 457 pages; ISBN 0 19 723056 3

No. 60
Burkitt's Lymphoma: A Human Cancer Model
Edited by G. Lenoir, G. O'Conor and C.L.M. Olweny
1985; 484 pages; ISBN 0 19 723057 1

No. 61
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Haloethers
Edited by M. Castegnaro, M. Alvarez, M. Iovu, E.B. Sansone, G.M. Telling and D.T. Williams
1985; 55 pages; ISBN 0 19 723061 X

No. 62
Directory of On-going Research in Cancer Epidemiology 1984
Edited by C.S. Muir and G. Wagner
1984; 717 pages; ISBN 0 19 723062 8
(out of print)

No. 63
Virus-associated Cancers in Africa
Edited by A.O. Williams, G.T. O'Conor, G.B. de Thé and C.A. Johnson
1984; 773 pages; ISBN 0 19 723063 6

No. 64
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Aromatic Amines and 4-Nitrobiphenyl
Edited by M. Castegnaro, J. Barek, J. Dennis, G. Ellen, M. Klibanov, M. Lafontaine, R. Mitchum, P. van Roosmalen, E.B. Sansone, L.A. Sternson and M. Vahl
1985; 84 pages; ISBN: 92 832 1164 2

No. 65
Interpretation of Negative Epidemiological Evidence for Carcinogenicity
Edited by N.J. Wald and R. Doll
1985; 232 pages; ISBN 92 832 1165 0

No. 66
The Role of the Registry in Cancer Control
Edited by D.M. Parkin, G. Wagner and C.S. Muir
1985; 152 pages; ISBN 92 832 0166 3

No. 67
Transformation Assay of Established Cell Lines: Mechanisms and Application
Edited by T. Kakunaga and H. Yamasaki
1985; 225 pages; ISBN 92 832 1167 7

No. 68
Environmental Carcinogens: Selected Methods of Analysis. Volume 7: Some Volatile Halogenated Hydrocarbons
Edited by L. Fishbein and I.K. O'Neill
1985; 479 pages; ISBN 92 832 1168 5

No. 69
Directory of On-going Research in Cancer Epidemiology 1985
Edited by C.S. Muir and G. Wagner
1985; 745 pages; ISBN 92 823 1169 3
(out of print)

No. 70
The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis
Edited by B. Singer and H. Bartsch
1986; 467 pages; ISBN 92 832 1170 7

No. 71
Environmental Carcinogens: Selected Methods of Analysis. Volume 8: Some Metals: As, Be, Cd, Cr, Ni, Pb, Se, Zn
Edited by I.K. O'Neill, P. Schuller and L. Fishbein
1986; 485 pages; ISBN 92 832 1171 5

No. 72
Atlas of Cancer in Scotland, 1975–1980: Incidence and Epidemiological Perspective
Edited by I. Kemp, P. Boyle, M. Smans and C.S. Muir
1985; 285 pages; ISBN 92 832 1172 3

No. 73
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Antineoplastic Agents
Edited by M. Castegnaro, J. Adams, M.A. Armour, J. Barek, J. Benvenuto, C. Confalonieri, U. Goff, G. Telling
1985; 163 pages; ISBN 92 832 1173 1

No. 74
Tobacco: A Major International Health Hazard
Edited by D. Zaridze and R. Peto
1986; 324 pages; ISBN 92 832 1174 X

No. 75
Cancer Occurrence in Developing Countries
Edited by D.M. Parkin
1986; 339 pages; ISBN 92 832 1175 8

No. 76
Screening for Cancer of the Uterine Cervix
Edited by M. Hakama, A.B. Miller and N.E. Day
1986; 315 pages; ISBN 92 832 1176 6

No. 77
Hexachlorobenzene: Proceedings of an International Symposium
Edited by C.R. Morris and J.R.P. Cabral
1986; 668 pages; ISBN 92 832 1177 4

No. 78
Carcinogenicity of Alkylating Cytostatic Drugs
Edited by D. Schmähl and J.M. Kaldor
1986; 337 pages; ISBN 92 832 1178 2

No. 79
Statistical Methods in Cancer Research. Volume III: The Design and Analysis of Long-term Animal Experiments
By J.J. Gart, D. Krewski, P.N. Lee, R.E. Tarone and J. Wahrendorf
1986; 213 pages; ISBN 92 832 1179 0

No. 80
Directory of On-going Research in Cancer Epidemiology 1986
Edited by C.S. Muir and G. Wagner
1986; 805 pages; ISBN 92 832 1180 4
(out of print)

No. 81
Environmental Carcinogens: Methods of Analysis and Exposure Measurement. Volume 9: Passive Smoking
Edited by I.K. O'Neill, K.D. Brunnemann, B. Dodet and D. Hoffmann
1987; 383 pages; ISBN 92 832 1181 2

No. 82
Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies
By N.E. Breslow and N.E. Day
1987; 404 pages; ISBN 92 832 0182 5

No. 83
Long-term and Short-term Assays for Carcinogens: A Critical Appraisal
Edited by R. Montesano, H. Bartsch, H. Vainio, J. Wilbourn and H. Yamasaki
1986; 575 pages; ISBN 92 832 1183 9

No. 84
The Relevance of *N*-Nitroso Compounds to Human Cancer: Exposure and Mechanisms
Edited by H. Bartsch, I.K. O'Neill and R. Schulte-Hermann
1987; 671 pages; ISBN 92 832 1184 7

No. 85
Environmental Carcinogens: Methods of Analysis and Exposure Measurement. Volume 10: Benzene and Alkylated Benzenes
Edited by L. Fishbein and I.K. O'Neill
1988; 327 pages; ISBN 92 832 1185 5

No. 86
Directory of On-going Research in Cancer Epidemiology 1987
Edited by D.M. Parkin and J. Wahrendorf
1987; 685 pages; ISBN: 92 832 1186 3
(out of print)

No. 87
International Incidence of Childhood Cancer
Edited by D.M. Parkin, C.A. Stiller, C.A. Bieber, G.J. Draper, B. Terracini and J.L. Young
1988; 401 page; ISBN 92 832 1187 1
(out of print)

No. 88
Cancer Incidence in Five Continents, Volume V
Edited by C. Muir, J. Waterhouse, T. Mack, J. Powell and S. Whelan
1987; 1004 pages; ISBN 92 832 1188 X

No. 89
Methods for Detecting DNA Damaging Agents in Humans: Applications in Cancer Epidemiology and Prevention
Edited by H. Bartsch, K. Hemminki and I.K. O'Neill
1988; 518 pages; ISBN 92 832 1189 8
(out of print)

No. 90
Non-occupational Exposure to Mineral Fibres
Edited by J. Bignon, J. Peto and R. Saracci
1989; 500 pages; ISBN 92 832 1190 1

No. 91
Trends in Cancer Incidence in Singapore 1968–1982
Edited by H.P. Lee, N.E. Day and K. Shanmugaratnam
1988; 160 pages; ISBN 92 832 1191 X

No. 92
Cell Differentiation, Genes and Cancer
Edited by T. Kakunaga, T. Sugimura, L. Tomatis and H. Yamasaki
1988; 204 pages; ISBN 92 832 1192 8

No. 93
Directory of On-going Research in Cancer Epidemiology 1988
Edited by M. Coleman and J. Wahrendorf
1988; 662 pages; ISBN 92 832 1193 6
(out of print)

No. 94
Human Papillomavirus and Cervical Cancer
Edited by N. Muñoz, F.X. Bosch and O.M. Jensen
1989; 154 pages; ISBN 92 832 1194 4

No. 95
Cancer Registration: Principles and Methods
Edited by O.M. Jensen, D.M. Parkin, R. MacLennan, C.S. Muir and R. Skeet
1991; 296 pages; ISBN 92 832 1195 2

No. 96
Perinatal and Multigeneration Carcinogenesis
Edited by N.P. Napalkov, J.M. Rice, L. Tomatis and H. Yamasaki
1989; 436 pages; ISBN 92 832 1196 0

No. 97
Occupational Exposure to Silica and Cancer Risk
Edited by L. Simonato, A.C. Fletcher, R. Saracci and T. Thomas
1990; 124 pages; ISBN 92 832 1197 9

No. 98
Cancer Incidence in Jewish Migrants to Israel, 1961-1981
Edited by R. Steinitz, D.M. Parkin, J.L. Young, C.A. Bieber and L. Katz
1989; 320 pages; ISBN 92 832 1198 7

No. 99
Pathology of Tumours in Laboratory Animals, Second Edition, Volume 1, Tumours of the Rat
Edited by V.S. Turusov and U. Mohr
1990; 740 pages; ISBN 92 832 1199 5
For Volumes 2 and 3 (Tumours of the Mouse and Tumours of the Hamster), see IARC Scientific Publications Nos. 111 and 126.

No. 100
Cancer: Causes, Occurrence and Control
Editor-in-Chief: L. Tomatis
1990; 352 pages; ISBN 92 832 0110 8

No. 101
Directory of On-going Research in Cancer Epidemiology 1989–1990
Edited by M. Coleman and J. Wahrendorf
1989; 828 pages; ISBN 92 832 2101 X

No. 102
Patterns of Cancer in Five Continents
Edited by S.L. Whelan, D.M. Parkin and E. Masuyer
1990; 160 pages; ISBN 92 832 2102 8

No. 103
Evaluating Effectiveness of Primary Prevention of Cancer
Edited by M. Hakama, V. Beral, J.W. Cullen and D.M. Parkin
1990; 206 pages; ISBN 92 832 2103 6

No. 104
Complex Mixtures and Cancer Risk
Edited by H. Vainio, M. Sorsa and A.J. McMichael
1990; 441 pages; ISBN 92 832 2104 4

No. 105
Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins
Edited by I.K. O'Neill, J. Chen and H. Bartsch
1991; 614 pages; ISBN 92 832 2105 2

No. 106
Atlas of Cancer Incidence in the Former German Democratic Republic
Edited by W.H. Mehnert, M. Smans, C.S. Muir, M. Möhner and D. Schön
1992; 384 pages; ISBN 92 832 2106 0

No. 107
Atlas of Cancer Mortality in the European Economic Community
Edited by M. Smans, C. Muir and P. Boyle
1992; 213 pages + 44 coloured maps; ISBN 92 832 2107 9

No. 108
Environmental Carcinogens: Methods of Analysis and Exposure Measurement. Volume 11: Polychlorinated Dioxins and Dibenzofurans
Edited by C. Rappe, H.R. Buser, B. Dodet and I.K. O'Neill
1991; 400 pages; ISBN 92 832 2108 7

No. 109
Environmental Carcinogens: Methods of Analysis and Exposure Measurement. Volume 12: Indoor Air
Edited by B. Seifert, H. van de Wiel, B. Dodet and I.K. O'Neill
1993; 385 pages; ISBN 92 832 2109 5

No. 110
Directory of On-going Research in Cancer Epidemiology 1991
Edited by M.P. Coleman and J. Wahrendorf
1991; 753 pages; ISBN 92 832 2110 9

No. 111
Pathology of Tumours in Laboratory Animals, Second Edition. Volume 2: Tumours of the Mouse
Edited by V. Turusov and U. Mohr
1994; 800 pages; ISBN 92 832 2111 1

No. 112
Autopsy in Epidemiology and Medical Research
Edited by E. Riboli and M. Delendi
1991; 288 pages; ISBN 92 832 2112 5

No. 113
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Mycotoxins
Edited by M. Castegnaro, J. Barek, J.M. Frémy, M. Lafontaine, M. Miraglia, E.B. Sansone and G.M. Telling
1991; 63 pages; ISBN 92 832 2113 3

No. 114
Laboratory Decontamination and Destruction of Carcinogens in Laboratory Wastes: Some Polycyclic Heterocyclic Hydrocarbons
Edited by M. Castegnaro, J. Barek, J. Jacob, U. Kirso, M. Lafontaine, E.B. Sansone, G.M. Telling and T. Vu Duc
1991; 50 pages; ISBN 92 832 2114 1

No. 115
Mycotoxins, Endemic Nephropathy and Urinary Tract Tumours
Edited by M. Castegnaro, R. Plestina, G. Dirheimer, I.N. Chernozemsky and H. Bartsch
1991; 340 pages; ISBN 92 832 2115 X

No. 116
Mechanisms of Carcinogenesis in Risk Identification
Edited by H. Vainio, P. Magee, D. McGregor and A.J. McMichael
1992; 615 pages; ISBN 92 832 2116 8

No. 117
Directory of On-going Research in Cancer Epidemiology 1992
Edited by M. Coleman, E. Demaret and J. Wahrendorf
1992; 773 pages; ISBN 92 832 2117 6

No. 118
Cadmium in the Human Environment: Toxicity and Carcinogenicity
Edited by G.F. Nordberg, R.F.M. Herber and L. Alessio
1992; 470 pages; ISBN 92 832 2118 4

No. 119
The Epidemiology of Cervical Cancer and Human Papillomavirus
Edited by N. Muñoz, F.X. Bosch, K.V. Shah and A. Meheus
1992; 288 pages; ISBN 92 832 2119 2

No. 120
Cancer Incidence in Five Continents, Vol. VI
Edited by D.M. Parkin, C.S. Muir, S.L. Whelan, Y.T. Gao, J. Ferlay and J. Powell
1992; 1020 pages; ISBN 92 832 2120 6

No. 121
Time Trends in Cancer Incidence and Mortality
By M. Coleman, J. Estéve, P. Damiecki, A. Arslan and H. Renard
1993; 820 pages; ISBN 92 832 2121 4

No. 122
International Classification of Rodent Tumours.
Part I. The Rat
Editor-in-Chief: U. Mohr
1992–1996; 10 fascicles of 60–100 pages; ISBN 92 832 2122 2

No. 123
Cancer in Italian Migrant Populations
Edited by M. Geddes, D.M. Parkin, M. Khlat, D. Balzi and E. Buiatti
1993; 292 pages; ISBN 92 832 2123 0

No. 124
Postlabelling Methods for the Detection of DNA Damage
Edited by D.H. Phillips, M. Castegnaro and H. Bartsch
1993; 392 pages; ISBN 92 832 2124 9

No. 125
DNA Adducts: Identification and Biological Significance
Edited by K. Hemminki, A. Dipple, D.E.G. Shuker, F.F. Kadlubar, D. Segerbäck and H. Bartsch
1994; 478 pages; ISBN 92 832 2125 7

No. 126
Pathology of Tumours in Laboratory Animals, Second Edition. Volume 3: Tumours of the Hamster
Edited by V. Turosov and U. Mohr
1996; 464 pages; ISBN 92 832 2126 5

No. 127
Butadiene and Styrene: Assessment of Health Hazards
Edited by M. Sorsa, K. Peltonen, H. Vainio and K. Hemminki
1993; 412 pages; ISBN 92 832 2127 3

No. 128
Statistical Methods in Cancer Research. Volume IV. Descriptive Epidemiology
By J. Estève, E. Benhamou and L. Raymond
1994; 302 pages; ISBN 92 832 2128 1

No. 129
Occupational Cancer in Developing Countries
Edited by N. Pearce, E. Matos, H. Vainio, P. Boffetta and M. Kogevinas
1994; 191 pages; ISBN 92 832 2129 X

No. 130
Directory of On-going Research in Cancer Epidemiology 1994
Edited by R. Sankaranarayanan, J. Wahrendorf and E. Démaret
1994; 800 pages; ISBN 92 832 2130 3

No. 132
Survival of Cancer Patients in Europe: The EUROCARE Study
Edited by F. Berrino, M. Sant, A. Verdecchia, R. Capocaccia, T. Hakulinen and J. Estève
1995; 463 pages; ISBN 92 832 2132 X

No. 134
Atlas of Cancer Mortality in Central Europe
W. Zatonski, J. Estéve, M. Smans, J. Tyczynski and P. Boyle
1996; 300 pages; ISBN 92 832 2134 6

No. 135
Methods for Investigating Localized Clustering of Disease
Edited by F.E. Alexander and P. Boyle
1996; 235 pages; ISBN 92 832 2135 4

No. 136
Chemoprevention in Cancer Control
Edited by M. Hakama, V. Beral, E. Bulatti, J. Faivre and D.M. Parkin
1996; 160 pages; ISBN 92 832 2136 2

No. 137
Directory of On-going Research in Cancer Epidemiology 1996
Edited by R. Sankaranarayan, J. Warendorf and E. Démaret
1996; 810 pages; ISBN 92 832 2137 0

No. 138
Social Inequalities and Cancer
Edited by M. Kogevinas, N. Pearce, M. Susser and P. Boffetta
1997; 412 pages; ISBN 92 832 2138 9

No. 139
Principles of Chemoprevention
Edited by B.W. Stewart, D. McGregor and P. Kleihues
1996; 358 pages; ISBN 92 832 2139 7

No. 140
Mechanisms of Fibre Carcinogenesis
Edited by A.B. Kane, P. Boffetta, R. Saracci and J.D. Wilbourn
1996; 135 pages; ISBN 92 832 2140 0

No. 142
Application of Biomarkers to Cancer Epidemiology
Edited by P. Toniolo, P. Boffetta, D. Shuker, N. Rothman, B. Hulka and N. Pearce
1997; 336 pages; ISBN 92 832 2142 7

No. 143
Cancer Incidence in Five Continents, Vol. VIII
Edited by D.M. Parkin, S.L. Whelan, J. Ferlay, L. Raymond and J. Young
1997; 1350 pages; ISBN 92 832 2143 5

IARC Technical Reports

No. 1
Cancer in Costa Rica
Edited by R. Sierra, R. Barrantes, G. Muñoz Leiva, D.M. Parkin, C.A. Bieber and N. Muñoz Calero
1988; 124 pages;
ISBN 92 832 1412 9

No. 2
SEARCH: A Computer Package to Assist the Statistical Analysis of Case-Control Studies
Edited by G.J. Macfarlane, P. Boyle and P. Maisonneuve
1991; 80 pages; ISBN 92 832 1413 7

No. 3
Cancer Registration in the European Economic Community
Edited by M.P. Coleman and E. Démaret
1988; 188 pages; ISBN 92 832 1414 5

No. 4
Diet, Hormones and Cancer: Methodological Issues for Prospective Studies
Edited by E. Riboli and R. Saracci
1988; 156 pages; ISBN 92 832 1415 3

No. 5
Cancer in the Philippines
Edited by A.V. Laudico, D. Esteban and D.M. Parkin
1989; 186 pages; ISBN 92 832 1416 1

No. 6
La genèse du Centre international de recherche sur le cancer
By R. Sohier and A.G.B. Sutherland
1990, 102 pages; ISBN 92 832 1418 8

No. 7
Epidémiologie du cancer dans les pays de langue latine
1990, 292 pages; ISBN 92 832 1419 6

No. 8
Comparative Study of Anti-smoking Legislation in Countries of the European Economic Community
By A. J. Sasco, P. Dalla-Vorgia and P. Van der Elst
1992; 82 pages; ISBN: 92 832 1421 8
Etude comparative des Législations de Contrôle du Tabagisme dans les Pays de la Communauté économique européenne
1995; 82 pages; ISBN 92 832 2402 7

No. 9
Epidémiologie du cancer dans les pays de langue latine
1991; 346 pages; ISBN 92 832 1423 4

No. 10
Manual for Cancer Registry Personnel
Edited by D. Esteban, S. Whelan, A. Laudico and D.M. Parkin
1995; 400 pages; ISBN 92 832 1424 2

No. 11
Nitroso Compounds: Biological Mechanisms, Exposures and Cancer Etiology
Edited by I. O'Neill and H. Bartsch
1992; 150 pages; ISBN 92 832 1425 X

No. 12
Epidémiologie du cancer dans les pays de langue latine
1992; 375 pages; ISBN 92 832 1426 9

No. 13
Health, Solar UV Radiation and Environmental Change
By A. Kricker, B.K. Armstrong, M.E. Jones and R.C. Burton
1993; 213 pages; ISBN 92 832 1427 7

No. 14
Epidémiologie du cancer dans les pays de langue latine
1993; 400 pages; ISBN 92 832 1428 5

No. 15
Cancer in the African Population of Bulawayo, Zimbabwe, 1963–1977
By M.E.G. Skinner, D.M. Parkin, A.P. Vizcaino and A. Ndhlovu
1993; 120 pages; ISBN 92 832 1429 3

No. 16
Cancer in Thailand 1984–1991
By V. Vatanasapt, N. Martin,

H. Sriplung, K. Chindavijak, S. Sontipong, S. Sriamporn, D.M. Parkin and J. Ferlay
1993; 164 pages; ISBN 92 832 1430 7

No. 18
Intervention Trials for Cancer Prevention
By E. Buiatti
1994; 52 pages; ISBN 92 832 1432 3

No. 19
Comparability and Quality Control in Cancer Registration
By D.M. Parkin, V.W. Chen, J. Ferlay, J. Galceran, H.H. Storm and S.L. Whelan
1994; 110 pages plus diskette; ISBN 92 832 1433 1

No. 20
Epidémiologie du cancer dans les pays de langue latine
1994; 346 pages; ISBN 92 832 1434 X

No. 21
ICD Conversion Programs for Cancer
By J. Ferlay
1994; 24 pages plus diskette; ISBN 92 832 1435 8

No. 22
Cancer in Tianjin
By Q.S. Wang, P. Boffetta, M. Kogevinas and D.M. Parkin
1994; 96 pages; ISBN 92 832 1433 1

No. 23
An Evaluation Programme for Cancer Preventive Agents
By Bernard W. Stewart
1995; 40 pages; ISBN 92 832 1438 2

No. 24
Peroxisome Proliferation and its Role in Carcinogenesis
1995; 85 pages; ISBN 92 832 1439 0

No. 25
Combined Analysis of Cancer Mortality in Nuclear Workers in Canada, the United Kingdom and the United States of America
By E. Cardis, E.S. Gilbert, L. Carpenter, G. Howe, I. Kato, J. Fix, L. Salmon, G. Cowper, B.K. Armstrong, V. Beral, A. Douglas, S.A. Fry, J. Kaldor, C. Lavé, P.G. Smith, G. Voelz and L. Wiggs
1995; 160 pages; ISBN 92 832 1440 4

No. 26
Mortalité par Cancer des Imigrés en France, 1979-1985
By C. Bouchardy, M. Khlat, P. Wanner and D.M. Parkin
1997; 150 pages; ISBN 92 832 2404 3

No. 27
Cancer in Three Generations of Young Israelis
By J. Iscovich and D.M. Parkin
1997; 150 pages; ISBN 92 832 2441 2

No. 29
International Classification of Childhood Cancer 1996
By E. Kramarova, C.A. Stiller, J. Ferlay, D.M. Parkin, G.J. Draper, J. Michaelis, J. Neglia and S. Qurechi
1996; 48 pages + diskette; ISBN 92 832 1443 9

IARC CancerBase

No. 1
EUCAN90: Cancer in the European Union (Electronic Database with Graphic Display)
By J. Ferlay, R.J. Black, P. Pisani, M.T. Valdivieso and D.M. Parkin
1996; Computer software on 3.5" IBM diskette + user's guide (50 pages); ISBN 92 832 1450 1

No. 2
CI5VII Electronic Database of Cancer Incidence in Five Continents. Vol. II
By J. Ferlay, R.J. Black, S.L. Whelan and D.M. Parkin
1997; Computer software on 3.5" IBM diskettes + user's guide (c. 48 pages); ISBN 92 832 1449 8

All IARC Publications are available directly from IARCPress, 150 Cours Albert Thomas, F-69372 Lyon cedex 08, France (Fax: +33 4 72 73 83 02; E-mail: press@iarc.fr).

IARC Monographs and Technical Reports are also available from the World Health Organization Distribution and Sales, CH-1211 Geneva 27 (Fax: +41 22 791 4857) and from WHO Sales Agents worldwide.

IARC Scientific Publications are also available from Oxford University Press, Walton Street, Oxford, UK OX2 6DP (Fax: +44 1865 267782).