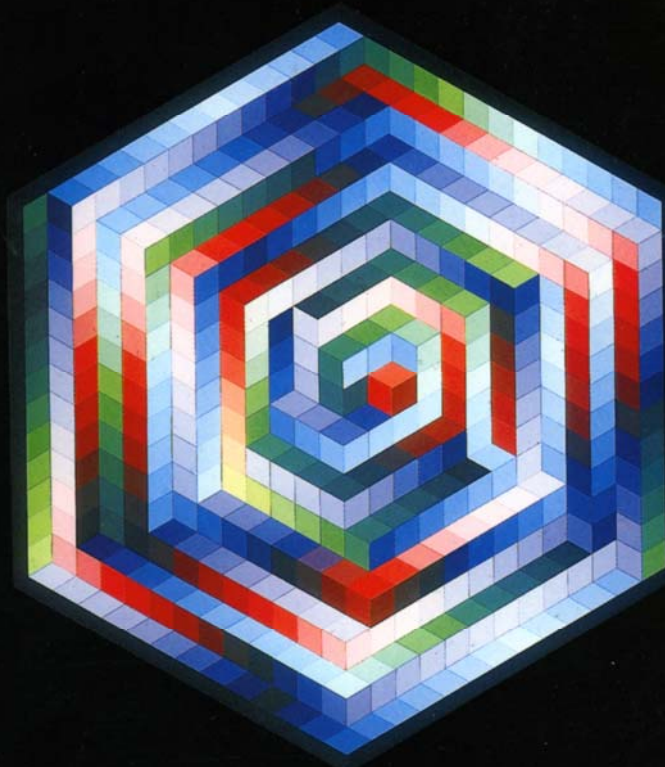


INTERNATIONAL AGENCY FOR RESEARCH ON CANCER (WHO)



STATISTICAL METHODS IN CANCER RESEARCH  
VOLUME IV

# DESCRIPTIVE EPIDEMIOLOGY



J. ESTEVE, E. BENHAMOU and L. RAYMOND  
Translated from French by M. Sinclair

IARC SCIENTIFIC PUBLICATIONS

N° 128  
LYON 1994

WORLD HEALTH ORGANIZATION



INTERNATIONAL AGENCY FOR RESEARCH ON CANCER

**Statistical Methods  
in Cancer Research  
Volume IV  
Descriptive Epidemiology**

**Jacques Estève, Ellen Benhamou, Luc Raymond**

*Translated from French by Mary Sinclair*

IARC Scientific Publications No.128

International Agency for Research on Cancer

Lyon

1994

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

---

© International Agency for Research on Cancer, 1994

Distributed by Oxford University Press, Walton Street, Oxford OX2 6DP, UK

Distributed in the USA by Oxford University Press, New York

First published 1993 in French under the title *Méthodes statistiques en épidé-*

*miologie descriptive* by Les Editions INSERM, Paris.

Publications of the World Health Organization enjoy copyright protection in accordance with the provisions of Protocol 2 of the Universal Copyright Convention.

All rights reserved.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

The authors alone are responsible for the views expressed in this publication.

The International Agency for Research on Cancer welcomes requests for permission to reproduce or translate its publications, in part or in full. Applications and enquiries should be addressed to the Editorial & Publications Service, International Agency for Research on Cancer, which will be glad to provide the latest information on any changes made to the text, plans for new editions, and reprints and translations already available.

### **IARC Library Cataloguing in Publication Data**

Statistical methods in cancer research /

Jacques Esteve, Ellen Benhamou, Luc Raymond

Contents: v. 4. Descriptive epidemiology

(IARC scientific publications; 128)

1. Neoplasms epidemiology 2. Statistics

I. Esteve, Jacques II. Benhamou, Ellen III. Raymond, Luc

IV. Series

ISBN 92 832 2128 1 (NLM Classification: QZ 206)

ISSN 0300-5085

Printed in France

**Statistical Methods  
in Cancer Research  
Volume IV  
Descriptive Epidemiology**

# 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 at 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 are intended to contribute to the dissemination of authoritative information on different aspects of cancer research. A complete list is printed at the back of this book.

*Cover illustration* : Victor Vasarely, HAT-1V, 1971

*Copyright* : © ADAGP, Paris, 1994

## Authors

### **Jacques Esteve**

Adviser on Biostatistics, International Agency for Research on Cancer

### **Ellen Benhamou**

Epidemiologist, Gustave Roussy Institute, Paris, France

### **Luc Raymond**

Epidemiologist, Geneva Cancer Registry and Department of Social and Preventive Medicine, University of Geneva, Switzerland

## Translator

### **Mary Sinclair**

Scientific Editor, Alpha Biomedical Communications, Sydney, Australia

# Contents

<b>Preface</b> .....	IX
<b>Foreword</b> .....	XI
<b>Acknowledgements</b> .....	XIII
<b>1. Fundamental concepts</b> .....	1
<b>Introduction</b> .....	1
<b>Basic concepts of descriptive epidemiology</b> .....	3
Time and the concept of incidence .....	3
Group characteristics and place .....	6
<b>Statistical concepts for the analysis of incidence data</b> .....	11
Formal definition of the incidence rate .....	11
Estimation of the instantaneous incidence rate .....	14
<i>An approximation useful in descriptive epidemiology</i> .....	14
<i>When individual observations are available and complete</i> .....	15
<i>When individual observations are available but possibly censored</i> .....	18
<b>Statistical concepts in survival analysis</b> .....	21
Follow-up studies .....	21
Survival probability .....	23
The life table .....	26
Classical models for survival distribution .....	29
<b>Interactive risks</b> .....	34
Competing risks .....	34
Relationship between incidence, mortality, survival and prevalence .....	37
<b>Bibliographical notes</b> .....	44
<b>References</b> .....	45
<b>2. Techniques for the analysis of cancer risk</b> .....	49
<b>Measurement of the risk of cancer</b> .....	49
Age- and sex-specific rates .....	49
<i>The calculation of a specific rate</i> .....	49
<i>The accuracy of the estimate of a rate</i> .....	51
<i>The incidence curve</i> .....	54
Standardized rates .....	56
<i>Direct standardization</i> .....	56
<i>Cumulative rates</i> .....	60
<i>Indirect standardization</i> .....	62
Probability of developing a specific form of cancer .....	66
The number of years of life lost .....	69
<i>Years of life lost with respect to a fixed age limit</i> .....	70
<i>Years of life lost with respect to life expectancy</i> .....	71

<b>Methods for comparison</b> .....	73
Comparison of incidence of a disease in two groups .....	74
<i>The approximate method</i> .....	74
<i>Mantel-Haenszel test</i> .....	77
<i>Overall measure of incidence ratio</i> .....	79
<i>Test of a multiplicative model</i> .....	81
<i>Trend test</i> .....	82
<i>Example</i> .....	83
Comparison of incidence among several populations .....	85
<i>Comparison with an overall expected value</i> .....	85
<i>Homogeneity test for incidence</i> .....	87
<i>Use of the log-linear model</i> .....	90
<b>Extension and limitations of the present methodology</b> .....	95
Risk analyses in the absence of denominators .....	95
<i>Standardized indices of relative frequency</i> .....	96
<i>Modelling incidence data in the absence of the denominator</i> .....	96
Choosing between various risk measures .....	98
<b>Bibliographical notes</b> .....	103
<b>References</b> .....	104
<b>3. Space-time variations and group correlations</b> .....	107
<b>Geographical analysis</b> .....	107
The objectives of cartography .....	107
Methods .....	111
<i>Geographical division</i> .....	111
<i>Choice of a risk indicator</i> .....	114
<i>Definition of risk classes</i> .....	114
Tools to interpret geographical data .....	119
<i>Autocorrelations</i> .....	119
<i>Identifying risk clusters</i> .....	122
<i>Time-space clustering</i> .....	131
<i>Smoothing and the empirical Bayes method</i> .....	134
<i>Concluding remark</i> .....	140
<b>Ecological studies</b> .....	141
Aim and methodological principles .....	141
Strengths and limitations of a measure of group exposure .....	148
<i>Group versus individual exposure</i> .....	148
<i>Risk estimation in the context of an ecological study</i> .....	149
<i>The ecological fallacy</i> .....	150
Specific techniques and examples .....	154
<i>Definition of groups</i> .....	154
<i>Multivariate analysis</i> .....	158
<i>Migrant studies</i> .....	166
<b>Time trends</b> .....	170
Objectives .....	170
Methods .....	172

<i>Components of temporal evolution</i> .....	172
<i>Effect of data quality</i> .....	176
<i>Role of modelling</i> .....	178
<i>Description of trend by period</i> .....	179
<i>Description of trend by cohort</i> .....	189
<i>Age-period-cohort models</i> .....	194
Specific techniques and examples .....	198
<i>Epidemiological evaluation of a secondary prevention campaign</i> .....	198
<i>Trends in cancer of the uterine cervix</i> .....	202
<b>Bibliographical notes</b> .....	204
<b>References</b> .....	206
<b>4. Techniques for survival analysis</b> .....	214
<b>Survival analysis in descriptive epidemiology</b> .....	214
<b>Estimation of survival distribution</b> .....	216
Estimation of crude survival .....	216
<i>Actuarial method</i> .....	216
<i>Kaplan-Meier method</i> .....	219
<i>Confidence interval for a survival rate</i> .....	222
<i>Median survival time</i> .....	225
<i>Collecting data for survival analysis</i> .....	227
Estimation of net survival .....	229
<i>Cause-specific survival</i> .....	230
<i>Relative survival</i> .....	231
<i>Constructing a life table</i> .....	236
<i>Alternative methods</i> .....	237
<b>Methods of comparison</b> .....	245
Introduction .....	245
Comparison of crude survival probabilities .....	246
<i>Test to compare two survival rates</i> .....	246
<i>Rank tests to compare survival curves</i> .....	247
Stratified comparison of crude survival .....	255
The Cox model .....	260
<i>The proportional hazard model</i> .....	260
<i>Principle of the Cox model estimation</i> .....	263
<i>The likelihood ratio test</i> .....	266
<i>Example of an adjustment using Cox's model</i> .....	268
Comparison of net survival .....	272
<b>Bibliographical notes</b> .....	276
<b>References</b> .....	277
<b>Appendix 1. Life table for Switzerland; 1978-1983</b> .....	281
<b>Appendix 2. Using GLIM in descriptive epidemiology</b> .....	284



## Preface

As the authors of this book remind us in their introduction, classical descriptive epidemiology was long regarded as simply a first, rather crude, step in the exploration of an epidemiological problem. Based essentially on comparisons between populations, it could do no more than stimulate ideas and hypotheses. Instead, it was up to analytical epidemiology, a more precise science since it involves measurements at the individual level, to produce firm evidence on risk factors, and it is mainly on the methodology of this area that various books have focused over the last thirty years.

Nevertheless, I can recall a number of major successes of descriptive epidemiology. For example, simply mapping the distribution of mortality rates for oesophageal cancer and of alcoholic cirrhosis enabled us, with Daniel Schwartz and Odile Lasserre, to demonstrate a relationship between alcohol consumption and oesophageal cancer. Likewise, comparison between the rising curve of lung cancer mortality and that of cigarette consumption certainly played a decisive role in focusing the attention of Richard Doll on the link between tobacco and cancer.

In recent years, the establishment of numerous cancer registries has encouraged many researchers to attempt to draw the maximum advantage from the data collected on cancer incidence and mortality. This has led to many original ideas and raised many important questions. Some biostatisticians entered this arena and they have gradually been able to lay the necessary statistical and mathematical foundations that were lacking. Problems such as those posed by the study of risks when the denominator is unknown, competing risks, and autocorrelation have led to the development of solid methodological concepts.

The great merit of this volume is that it brings together and reviews in one coherent text the different techniques needed for a modern approach to descriptive epidemiology. With the help of this compilation, researchers in this field will henceforth be able to tackle the study of their data armed with a methodological arsenal giving them the optimal chance of success. Other readers such as doctors and public health specialists will be able to obtain guidance, without having to enter into all the mathematical details, on how to avoid the many pitfalls that confront those who have to interpret collections of numerical data. The authors make extensive use of examples of analysis of real data sets and show how these can be treated and interpreted, so that the reader can follow in detail the development of the methods described and better comprehend the range of their applications.

I have known Jacques Estève, Ellen Benhamou and Luc Raymond for very many years, as a productive team with complementary capabilities. I am certain that this book that they have co-authored will provide an indispensable guide for numerous researchers and for decision-makers in public health who are concerned with epidemiology. Personally, as an epidemiologist and oncologist, I have found it to be of the greatest interest.

Professor Robert Flamant  
*Director of the Gustave Roussy Institute,  
Villejuif, France*

## Foreword

This monograph presents and discusses some methods used in descriptive epidemiology which are relevant to cancer research. In presenting the fundamental concepts, we have tried to keep the mathematical formulation at a level which is compatible with an elementary knowledge of statistics and probability, but which nevertheless enables the logical relationships between the concepts currently used in epidemiology to be understood.

With the above objective in mind. Chapter 1 describes briefly the epidemiological context in which the methods will be used and devotes some space to their mathematical formulation. An elementary knowledge of statistics and probability as well as some familiarity with mathematical reasoning is expected from the reader of this chapter.

Chapter 2 describes how, in practice, the analysis and comparison of incidence and mortality can be carried out. Most attention is given to the multiplicative model and to the concept of proportional hazards, which is particularly relevant to cancer research. The exposition of these notions relies on many numerical examples, but no great mathematical sophistication is needed.

Chapter 3 is devoted to geographical analysis, ecological studies and analysis of time trends. These fields are at present subjects of interesting methodological research, and we have tried to show from several examples how modern statistical tools can considerably improve the interpretation of geographical and temporal data in epidemiology.

Chapter 4 describes the methods of analysis of survival probability at an elementary mathematical level, and the emphasis is placed on the interpretation of such data when they are collected in the context of routine cancer registry operations. Much space is therefore given to the concept of relative survival and many examples are presented to show the difficulty of interpretation when the procedures for data collection may imply several types of bias.

We have tried to give the reader sufficient understanding to use the methods which are presented by giving the details of calculations whenever possible and some examples of the use of the GLIM software, which is cheap, widely available and enables many methods presented in this book to be readily implemented.

This text was first written in French and was translated by Mary Sinclair, whom we gratefully acknowledge for her careful work. With the exception of the correction of known errors and some inevitable adaptation of French to English style, no effort has been made to update the content which was essentially written before 1991; this is why some recent references which would have been relevant are not included.

The finalization of the manuscript of this monograph benefited from the careful reading of John Cheney. We gratefully acknowledge his help.

# Acknowledgements

This book was originally conceived as the proceedings of two seminars on statistical methods organized by the group of cancer registries in Latin countries. However, the absence in the French scientific literature of an up-to-date text on methods in descriptive epidemiology led us to write a manual on the key concepts and statistical methods in this area.

We hope that this manual will be of value to those working in, or with cancer registries, for whom it was originally intended. We trust that the mathematical nature of some parts of the book will not be an obstacle to its use and that the inclusion of numerous examples will serve to establish the link between the theoretical development and its epidemiological interpretation.

Many people have contributed to the development of this book: most notably the participants at the seminars mentioned above, whom we gratefully acknowledge. We are especially indebted to those who have made unpublished data available to illustrate the procedures presented in this book.

We are especially grateful to Catherine Com-Nougue and Agnes Laplanche who carefully read the manuscript and assisted us with pertinent advice and suggestions; to Annick Rivoire for her invaluable contribution in the preparation and production of the book; and to Aude Jaccard for organizing the bibliography. Finally, we thank Nicholas Day for his encouragement and help when initiating this project.

St Just et Vacquières, April 1992