Chapter 4. Information bias: misclassification and mismeasurement of exposure and outcome

CHAPTER 4.

Information bias: misclassification and mismeasurement of exposure and outcome

Leslie Stayner, Neil Pearce, Ellen Aagaard Nøhr, Laura Beane Freeman, Veronika Deffner, Pietro Ferrari, Laurence S. Freedman, Manolis Kogevinas, Hans Kromhout, Sarah Lewis, Richard MacLehose, Marie-Elise Parent, Lorenzo Richiardi, Pamela Shaw, and Roland Wedekind

4.1 Introduction

Nearly all epidemiological studies of carcinogenic hazards suffer to some degree from error due to the methods used to measure exposures and outcomes; this error is commonly referred to as measurement error or misclassification (described in this chapter; see also the Preface). Measurement error can occur both in studies that use continuous measures of exposure and in studies that use categorical measures. Any bias resulting from such error is generally referred to as information bias (Lash et al., 2021).

Exposure assessments based on questionnaires are often prone to several sources of measurement error. Of particular concern is the validity of exposure information from interviews of the next of kin rather than the study participants themselves. In occupational studies, exposure assessments are commonly based on the development of a job-exposure matrix (JEM), which assigns exposures to individuals on the basis of their job, department, industry, or time period (or a combination of these) (Stewart et al., 1996). This often introduces errors, because not everyone assigned to an exposure group is likely to have the same exposure.

Even in the rare instance that objective physical measurements are available to estimate individual exposures, there is still a potential for exposure measurement error due to the instrumentation used. For example, personal measurements of radiation exposure using radiation dosimeters have been used in numerous epidemiological studies. Exposure estimates used in these studies will be subject to measurement errors, which could vary with the different radiation dosimeters used over time (Daniels and Schubauer-Berigan, 2005; Stayner et al., 2007; Thierry-Chef et al., 2007, 2015).

Epidemiologists frequently use qualitative categories of potential exposure (e.g. high, medium, or low) when quantitative data on exposures are lacking, or to create categories from what is truly a continuous measure of exposure, using cut-points that may reflect the distribution of exposures in the study population (e.g. percentiles). Exposure misclassification occurs when study participants are incorrectly categorized with respect to their true exposure. Categorization can result in information bias due to mismeasurement of the individual exposures. In other words, an individual may have been placed in a high exposure group

who should have been placed in a lower exposure group, or vice versa. Misclassification may also occur in circumstances where the exposure is naturally categorical. For example, in some studies participants are classified as having ever been exposed or never been exposed. If this categorization is based on questionnaire data or inadequate work history information, then exposure misclassification may occur.

Measurement error and misclassification can be either differential or non-differential. Errors in exposure measurement or classification are differential when they vary by disease status. For example, differential misclassification of exposure may occur in a case–control study that uses questionnaire data collected after the outcome was observed. Case participants may be more likely than control participants to recall past exposures, because case participants may be searching for an explanation for their disease. This could result in case participants recalling their exposure more accurately than control participants, because they may have spent more time thinking about the possible causes of their disease. However, this could also mean that reporting of exposures by case participants is less accurate than that by control participants (e.g. if there is social stigma around the exposure and/or the outcome). This type of bias is called recall bias. Non-differential exposure measurement errors occur when the rate of misclassification is equal between participants in the case and control groups or, in other words, when the measurement error is independent of the disease status. For example, differential misclassification of exposure would be unlikely in a prospective cohort study, in which exposures are measured before follow-up, when the investigators had no information on future disease status.

The potential for misclassification or mismeasurement of exposure is particularly applicable to cancer studies, because the etiologically relevant exposures for most carcinogens are, in general, longer than the preceding 5–10 years, for leukaemias (Finkelstein, 2000; Schubauer-Berigan et al., 2007a, b), or the preceding 10–20 years, for solid tumours. Often, records of exposure measurements during the early years of a study do not exist or can only be estimated with a large degree of uncertainty. In many situations, historical measurements of exposure have been collected for regulatory compliance purposes and may be focused on documenting that the highest exposures are below occupational or environmental standards. Thus, historical measurements may not be representative of past exposures, and this could lead to substantial measurement error.

Misclassification of disease status can also be differential or non-differential with respect to exposure status. Non-differential misclassification occurs when there is overascertainment or underascertainment of disease, and the probability of disease misclassification is the same for exposed and unexposed study participants. Differential misclassification occurs when case identification is more accurate or less accurate in exposed participants than in unexposed participants. For example, women who work night shifts may be less likely to undergo breast cancer screening, and this may result in underdiagnosis (or late diagnosis) of breast cancer. In epidemiological studies of cancer risk, misclassification of disease is perhaps a less common issue than misclassification of exposure. However, there are exceptions, such as when studies of cancers with a low fatality rate are based on death certificate diagnosis rather than incident cases from tumour registries, or when data on outcomes are poorly recorded (e.g. in lower-income countries) or may simply be unavailable or of poor quality. Such misclassification would typically be non-differential with respect to exposure status.

In the past, epidemiologists and statisticians have perhaps paid insufficient attention to evaluating the potential for biases resulting from measurement error and misclassification of exposure or disease (Shaw et al., 2018). Non-differential exposure error typically creates a bias towards the null (i.e. towards observing no effect), but this is not always the situation, as discussed in Section 4.2.1. There has been an increasing trend in the development and use of new methods to assess the direction and magnitude of bias and to bias-adjust the effect measures to correct for measurement error (e.g. Cole et al., 2006; Lash et al., 2014; Corbin et al., 2017; Keogh et al., 2020; Shaw et al., 2020). In this chapter, we discuss these approaches with particular emphasis on methods that can be used with published studies to assess misclassification and measurement error in exposure and outcome, because *IARC Monographs* reviewers and other expert review groups would seldom have access to the raw data from epidemiological studies. We start by discussing

qualitative approaches for evaluating the direction of bias due to errors in exposure, considering first continuous and then categorical exposures.

4.2 Qualitative evaluation of the direction of bias due to errors in exposures

4.2.1 Non-differential errors in exposure

(a) Measurement errors of continuous variables

The direction of the bias associated with measurement errors of continuous exposures depends on which error models apply (see Side Box 4.1 for the definitions).

Classical non-differential measurement errors are expected to lead, on average, to underestimation of the association between the exposure and the disease. Thus, although the measurement method is itself unbiased, in the sense that the average measured exposure is equal to the true exposure, the estimated exposure–cancer association arising from such measurements tends to be biased towards the null value, on average (Spearman, 1904; Armstrong, 1998).

Under a linear model in which the measurements are not, on average, equal to the true value (i.e. are biased) and the measurement errors are non-differential, the bias can, theoretically, lead to either overestimation or underestimation of associations between an exposure and a health outcome. However, when a linear model is applied to self-reported dietary and physical activity data, the random errors are often so large that they dominate and, as with the classical model, lead, on average, to underestimation of exposure–cancer associations (Freedman et al., 2011).

Key message

Berkson errors are special and are different from classical errors in that they are not expected to appreciably distort the exposure–response relation, for example when the assigned exposures are the means of the true dose in the groups (Gilbert, 2009).

However, as in the classical error model, Berkson errors do reduce the precision of the estimated exposure–response relation.

In the event that Berkson errors are correlated with covariates in the outcome model, appreciable distortion of the exposure–response relation can result, and the association may be biased towards underestimation or overestimation in an unpredictable manner (see Keogh et al., 2020).

(b) Misclassification of categorical variables

The direction and magnitude of bias associated with non-differential misclassification of categorical exposure variables will depend on how many categories have been used, how accurate the assessment of the exposure is, and the prevalence of the exposure.

In a situation where a single exposure is declared present or absent, non-differential misclassification occurs when the sensitivity (the probability of having been identified as exposed when the individual is truly exposed) and the specificity (the probability of having been identified as unexposed when the individual is truly unexposed) of the errors are the same for cases and non-cases of disease.

Key message

Non-differential misclassification of a dichotomous exposure (exposed or unexposed) will, on average, result in attenuation of effect estimates towards the null (Wacholder, 1995; Armstrong, 1998), as seen in Example 4.3.

One should realize that any given study could still show a bias away from the null due to random variability, given that any study is simply a single realization of a measurement process and may deviate from the expectation (Jurek et al., 2005; Loken and Gelman, 2017). However, the larger the sample size, the smaller this chance (Wacholder, 1995; Yland et al., 2022).

Key message

The extent of the expected attenuation from non-differential exposure misclassification will depend on the prevalence of the exposure and the specificity and sensitivity of the exposure assessment and assignment.

When there are several categories (e.g. unexposed, low, medium, or high), non-differential misclassification can result in the overestimation of risk in an intermediate exposure category and the underestimation of risk in the highest category.

Misclassification might even change the direction of the slope across exposure categories (Dosemeci et al., 1990), unless the true exposure–response relation is positive and monotonic (Weinberg et al., 1994).

Side Box 4.1. Three common models describing measurement error in epidemiological studies

Besides the issue of whether the exposure measurement error is differential or non-differential, another aspect that influences the effect of the error on the results is the relation of the erroneous measurement to its underlying true value. This relation is usually described in terms of a statistical model. Any type of model is possible, but for continuous exposure variables (e.g. the time spent using a mobile phone over a specified period, or the mass of red meat consumed on a typical day), three models (described here) are most commonly found in the epidemiological literature. Because the impact (or non-impact) of the error on the estimated associations depends on the type of error, it is important for those reviewing the literature to know about them. These models all postulate additive random error. Multiplicative error can sometimes be handled by these models through transformation of the variables to a logarithmic scale. More-complex models involving random error that is "shared" between individuals have been postulated recently for occupational cohort studies (Stram and Kopecky, 2003; Hoffmann et al., 2018) but are not covered here.

(a) Classical model

This is the simplest model to describe measurement errors. If X denotes the true underlying exposure value and X^* denotes the measured value, then the relation between them is described by the model as

$$X^* = X + U \qquad \text{(E4.1)}$$

where U is a random error that has a mean of zero and is independent of the true value X. Thus, the model describes an erroneous measurement method that gives the correct value on average but yields a somewhat different value each time it is applied, sometimes larger than and sometimes smaller than the true exposure. Because the average error is zero, such a measurement method is called *unbiased*. Such measurements are commonly encountered in laboratory work, for example with assessments of serum levels of cholesterol (Glasziou et al., 2008) or C-reactive protein (Koenig et al., 2003). This model is also used when one is interested in an individual's average value of the measure over a specified period (the true value) but the measure is determined only once (or a few times) within the study period.

(b) Linear model

A somewhat more complex model is required for measurements that are not, on average, equal to the correct value. One way of describing such measurements, which is often used for self-reported dietary intake and physical activity data, is to postulate a linear relation between the measurement and its true value, as

$$X^* = \alpha_0 + \alpha_X X + U \qquad \text{(E4.2)}$$

where α_0 and α_X are the intercept and the slope, respectively, of the linear relation, and U, as before, is a random error that has a mean of zero and is independent of the true value X (see Keogh et al., 2020). The intercept α_0, known as the location bias, shifts the measurements up or down on average, while the slope α_X, known as the scale bias, governs how much the mismeasurement depends on the true value of the exposure. Although this model includes the classical model as a special case (when $\alpha_0 = 0$ and $\alpha_X = 1$), in its general form the model describes an erroneous measurement method that, on average, gives not the correct value X but an incorrect value $\alpha_0 + \alpha_X X$. Because of this property, such a measurement method is called *biased*. Such measurements are commonly encountered in self-reported behaviours (e.g. dietary intake). It is often found that α_0 is greater than 0 and α_X is positive but less than 1. Such values describe a pattern when underreporting becomes more severe as the true exposure increases (Example 4.1).

Note that, as in this example, the exposure is often measured on a logarithmic scale, and the additive random error becomes multiplicative on a linear scale.

Side Box 4.1. Three common models describing measurement error in epidemiological studies (continued)

Example 4.1. Linear models for measurement error of protein intake from food frequency questionnaires

Kipnis et al. (2003) used data from the Observing Protein and Energy Nutrition (OPEN) study and reported that for natural log-transformed self-reported total protein intake using a food frequency questionnaire, the value of α_X for men was 0.67. From the reported geometric mean intakes of protein in that study (Table 2 of Subar et al., 2003), one can calculate that α_0 was 1.18. These values imply that for a low total protein intake of 68.3 g/day (2.5th percentile), the average reported intake was exp[1.18 + 0.67ln(68.3)] = 55.1 g/day, with an underestimation of 19%, whereas for a high total protein intake of 158.3 g/day (97.5th percentile), the average reported intake was exp[1.18 + 0.67ln(158.3)] = 96.9 g/day, with a much larger degree of underestimation (39%).

(c) Berkson model

Another type of error, called Berkson error (Berkson, 1950), is only subtly different from the classical model but is important, both because it arises in many epidemiological settings and because its effects on results are very different from those of classical error. The relation between the measured value and the true value is described by this model as

$$X = X^* + U \qquad \text{(E4.3)}$$

where U is a random error that has a mean of zero and is independent of the measured value X^* but is not independent of the true value X. Berkson error commonly occurs in occupational health studies, when individual workers in the same job group are assigned the average measured exposure of their group or an exposure based on a JEM. In these cases, the true exposure of an individual equals the mean exposure in the job group to which the individual is assigned plus some independent random error. Berkson errors may also occur in studies of environmental exposures (Example 4.2). (text continues on page 90)

Example 4.2. Berkson error in an example from blood lead and intelligence quotient testing

In a study (Armstrong, 1998), the intelligence quotient measured at age 10 years of children living in the vicinity of a lead smelter was studied in relation to the children's exposure to lead. Blood lead levels were measured in a random sample of the study group; the full study group was then classified into subgroups according to the distances of their homes from the smelter, and the average blood lead level in each subgroup was assigned as the exposure level for all the children in that subgroup. Such an exposure measure can be assumed to have Berkson error, in the same way as for exposure assessments based on a JEM.

For the different impacts of classical errors, linear measurement errors, and Berkson errors, see Section 4.2.1(a).

Example 4.3. Non-differential exposure misclassification when exposure is rare versus when exposure is common

In a general population case–control study with a low prevalence (< 10%) of occupationally exposed individuals, low specificity will result in a large number of false-positives for the exposure and consequently result in considerable attenuation towards the null (Flegal et al., 1986). For this reason, when JEMs aim to assess occupational exposure in the general population where exposure is rare (e.g. population-based case–control studies), specificity should be favoured over sensitivity (Kromhout and Vermeulen, 2001). In contrast, in studies with a high prevalence of exposure (e.g. industrial cohort studies), low sensitivity will result in attenuation towards the null; therefore, sensitivity should be favoured over specificity. (text continues on page 90)

Key message

Misclassification of exposure may also occur when a continuous error-prone exposure variable (e.g. cumulative exposure) is categorized (Example 4.4). Categorization of a continuous exposure variable with error can actually result in differential misclassification if the probability of disease is a function of the continuous exposure rather than of the exposure categories (Flegal et al., 1986).

The expected magnitude of the bias in the intermediate categories will depend on how much the risk of disease differs across exposure groups and the actual shape of the exposure–response relation (Yland et al., 2022).

4.2.2 Exposure measurement errors that could be non-differential or differential: interviewer error or bias

In studies that involve an expert-based approach to assess exposures (e.g. having an expert panel of industrial hygienists assess exposures on the basis of work histories obtained by interview), the interviewer can play a critical role in obtaining the description of the tasks, agents, or protective measures that will be used to infer exposures. There is evidence that interview quality can lead to non-differential exposure misclassification and bias towards the null (Edwards et al., 1994), as in Example 4.5. Some interviewers can be more knowledgeable than others and elicit more clues; this will influence the reliability of the information (Example 4.6). Interviewer bias is also possible when additional information on exposure (e.g. asbestos exposure) is elicited by an interviewer who believes that asbestos is associated with the disease of the interviewee (e.g. lung cancer, mesothelioma), or the interviewer may not question control participants as deeply as case participants. These problems can, to some extent, be overcome by better interviewer training or by blinding interviewers to case–control status, although such blinding is rarely possible in cancer case–control studies (Edwards et al., 1994). These issues are addressed further in Section 4.2.4(b), in the context of negative control exposures.

4.2.3 Differential errors in exposure

Bias from differential errors in exposure can occur in both cohort and case–control studies. However, it is perhaps more common in case–control studies in which information on exposure is collected using

Example 4.4. Misclassification from categorizing a continuous exposure variable in workers exposed to crystalline silica

A pooled case–control study of respirable crystalline silica exposure and lung cancer (Ge et al., 2020) showed a largely flat exposure–response relation, particularly in the middle exposure categories (odds ratios [ORs] of 1.15, 1.33, 1.29, and 1.45 for cumulative exposure quintiles of > 0–0.39, 0.40–1.09, 1.10–2.39, and ≥ 2.40 mg/(m^3·years), respectively), whereas the analysis with continuous cumulative exposure showed a monotonic linear increase in risk for both untransformed and log-transformed exposure. (text continues above)

Example 4.5. Assessing for varying quality of the interviewee response in assessing tobacco smoking

Villanueva et al. (2009) conducted a multicentre hospital-based study of 1219 patients with incident bladder cancer and 1271 control participants, recruited in Spain in 1998–2001. Study information was obtained by trained interviewers, who administered structured computer-assisted personal interviews. The information was categorized into five sections (sociodemographic, smoking, occupational, residential, and medical history). At the end of each interview, the interviewer recorded the perceived quality of the interview for each section as unsatisfactory, questionable, reliable, or of high quality. It was found that 10% of the interviews were of unsatisfactory quality with regard to smoking history. It was also found that the strength of the association between cigarette smoking and bladder cancer increased with increasing interview quality, from an odds ratio of 3.20 (95% confidence interval [CI], 1.13–9.04) for interviews scored as unsatisfactory or questionable overall (taking into account all of the variables considered in the interviews) to an odds ratio of 7.70 (95% CI, 3.64–16.30) for high-quality interviews. Lower-quality interview scores were found with increasing age, poorer self-perception of health, and low socioeconomic status. However, differences were not found in the quality of interviews according to case or control status: 9% of patients had unsatisfactory or questionable interviews, compared with 7% of control participants (P = 0.109). (text continues on page 93)

Example 4.6. Assessing for varying quality of interviewer in assessing job histories

In a validity study, reports of job histories were compared with employers' records (Baumgarten et al., 1983). There was no evidence that the quality of job history information obtained from control participants was systematically different from that obtained from patients with cancer, although there was some evidence that different interviewers obtained job histories of varying quality, irrespective of case–control status. (text continues on page 93)

questionnaires administered retrospectively, after the disease under study has been diagnosed in the case participants.

Key message

Recall bias is not an inherent feature of case–control studies; for example, exposure estimation may be based on historical records (e.g. work history records) or biospecimens banked in the past.

When exposure assessment is based on objective measures, a case–control study is no more prone to information bias than the corresponding cohort study that uses the same exposure history records. However, many case–control studies do involve retrospective collection of exposure information; therefore, in this section, several types of differential information bias are considered that are of particular concern in case–control studies of this type.

(a) Recall and information bias

Case–control studies are often portrayed as being more prone to information bias when they involve the use of exposure questionnaires. This is not unique to case–control studies. Many cohort studies involve exposure questionnaires (on opium use, meat consumption, night shift work, etc.) at baseline and at follow-up. However, a potential additional problem in case–control studies is that exposure questionnaires are usually administered after the case or control status is known by the participants, and often also by the interviewers.

To understand the differential nature of this misclassification, consider that someone who has developed cancer is likely to have thought a great deal about the possible causes of their condition and may have sought further information (e.g. from the Internet). The same will usually not apply to control participants

drawn from the general population. For example, it has been suggested that patients with cancer may recall minor exposures to pesticides (e.g. spray drift from a neighbouring farm), whereas control participants from the general population may not recall such minor exposures (Smith et al., 1988). In this situation, differential recall could occur, and the proportion of case participants reporting past exposure to pesticides may be greater than the proportion of control participants, even if the pesticides actually do not cause the type of cancer under study. It is important to emphasize that such recall bias does not necessarily involve biased recall by the case participants; in fact, it may involve a lack of recall by the control participants. Examples 4.7 and 4.8 illustrate some of these important concepts surrounding recall bias with respect to two key topics.

(b) Differential information when provided by proxies

Proxies are sometimes recruited in studies of cancers with poor prognoses or of aggressive types of cancer, to better cover the base population of case participants. However, proxy respondents can sometimes provide information of a poorer quality than self-respondents; this can bias findings if the quality of exposure information differs by case status (Example 4.9).

Example 4.7. Recall bias and knowledge of carcinogenicity

Most studies of shift work are based on self-reported information about current and previous jobs. Information on job history and periods of work has been repeatedly shown to be accurately recalled. Recall of shift work details of previous jobs is more complex and may be prone to exposure misclassification. For example, in a case–control study in Spain (MCC-Spain), the frequency of shift work (nights per month) was more difficult to recall than its duration, and this led to a higher proportion of missing data (Papantoniou et al., 2016). It is unlikely that differential recall has been important in case–control studies of shift work and cancer. The potential carcinogenicity of night shift work was not well known in the wider population in the past 10–20 years, when most existing studies were conducted. However, recall bias is not necessarily avoided for this reason if night shift workers report differentially on factors that could be intermediate factors associated with disease, such as sleep. There do not seem to be any published studies examining this type of differential recall in detail. (text continues above)

Example 4.8. Estimation of the extent of recall bias

In the Interphone study (Vrijheid et al., 2009), validation studies were conducted to assess the potential for differential misclassification of self-reported mobile phone use. The investigators collected mobile phone records of case and control participants from network operators in three countries over an average of 2 years and compared them with self-reported mobile phone use. The ratio of reported to recorded phone use was estimated. Mean ratios were very similar for case and control participants; both underestimated the number of calls (mean ratio, 0.81) and overestimated call duration (mean ratio, 1.4). For case participants, but not control participants, the ratios were further away from 1.0 for time periods further before the interview. In addition, the ratios were greater for higher levels of use. These findings are very provisional, because they were based on records obtained for only a few participants with the relevant data. Nevertheless, based on the available data, there was little evidence for differential recall errors overall or in recent time periods. In contrast, there appeared to be overestimation of use by case participants in more distant time periods; this could cause positive bias in estimates of the odds ratios for mobile phone use. (text continues above)

Example 4.9. Proxy respondents and recall bias in a study of pesticide exposure

Brown et al. (1991) conducted a methodological study to compare information on pesticide use from farmers and their surrogates. The study included 95 farmers and their spouses or other close family members. Both the farmers and the proxies were asked about the farmers' pesticide use. Although there was good agreement between the farmer and the proxy about whether seven common pesticides had ever been used, there was much more variable agreement between the two regarding the frequency of use, with correlation coefficients ranging from 0.23 to 0.80 for number of days of use.

Later, the same researchers recruited proxy respondents in a series of case–control studies focused on pesticides and non-Hodgkin lymphoma. In a publication focused on the risk of non-Hodgkin lymphoma and use of the insecticide lindane, Blair et al. (1998) evaluated the effect of information provided by next-of-kin proxy respondents on risk estimates. Both living and deceased people were included, and control participants for deceased people in the case group were identified from death records and matched on age and year of death. For these deceased people, interviews were conducted with their next of kin, and living participants provided information directly. Study participants who could not recall whether they (or their proxies) had used lindane were excluded from analysis. The percentage of living case participants who could not recall whether they had used lindane was 6.0%, while that for proxy respondents of deceased people was 8.2%; 9.6% of living control participants and 11.1% of proxy respondents of deceased control participants could not recall whether lindane had been used. In addition, results were stratified by whether information on lindane was provided directly by the case or control participant or by a proxy. The odds ratio for whether lindane had ever been used was 1.3 (95% CI, 0.9–1.8) for direct respondents and 2.1 (95% CI, 1.0–4.4) when information was provided by a proxy. Similar differences in risk were seen for the number of days of use of lindane and whether or not personal protective equipment was used during application, with higher associations among those with information provided by a proxy. Although other factors could explain these results, differential misclassification of exposure could not be ruled out. (text continues below)

4.2.4 Tools for assessing differential exposure information bias

When a published paper is considered, it is important to assess the potential for information bias, as well as its probable magnitude and direction. A key issue is whether any misclassification of (categorical) exposure or disease is likely to be non-differential or differential. This section is particularly focused on the situation where information bias is likely to be differential, although many of the methods can also be used to assess non-differential information bias. We particularly consider assessment using substantive knowledge (external to the published paper) and the use of directed acyclic graphs (DAGs; see Chapter 2). As in Chapter 3, some tools are outlined that expert review groups can use to examine the influence of exposure measurement error.

(a) Tool E-1: use of substantive knowledge and DAGs for misclassification

Assessing the potential for differential information bias requires expert knowledge, usually from previously published studies, and mechanistic knowledge. The key feature of differential information bias is that the misclassification of exposure depends on disease status, or vice versa (the misclassification of disease status depends on exposure). For differential misclassification of (categorical) exposure status, this means that the sensitivity or specificity (or both) of the exposure measurement instrument is different for those with or without disease.

Misclassification can be summarized using a DAG (Hernán and Cole, 2009); these are covered in detail in Chapter 2 and are only briefly considered here. A DAG can help to clarify whether disease or exposure misclassification is differential or non-differential, for example when people with cancer (case participants) are likely to have different recall of past exposures

compared with healthy control participants. A similar bias can occur when there is a factor (e.g. ethnicity, socioeconomic status) that is a risk factor for disease (e.g. the disease is more common among less-affluent people) and affects the accuracy of exposure recall (e.g. less-affluent people are less aware of, or have different recall of, past exposures). Researchers can use DAGs to help determine whether differential misclassification, through a variety of mechanisms, is plausible.

The DAG will not identify whether such a bias is likely to occur or its probable magnitude and direction, but it does provide a framework for considering whether such a bias is possible and assessing any strategies that the investigators may have adopted to minimize, to control for, or to assess it (Example 4.10).

The use of DAGs can help study reviewers to identify whether differential or non-differential bias is possible in a given study. When several different studies are conducted for the same exposure–outcome relation, it is important to note that the DAG could be different for each study; some studies may be more or less prone to differential or non-differential misclassification, depending on the study design.

(b) Tool E-2: negative control exposures and positive control outcomes

A negative control exposure approach involves assessing the association with another exposure that is not associated with the outcome under study but is likely to be subject to a similar information bias (Lipsitch et al., 2010; Arnold et al., 2016; Lawlor et al., 2016).

Key message

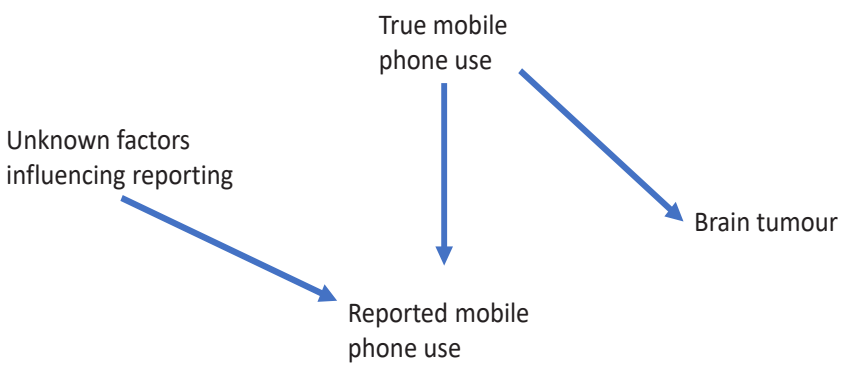
A key assumption of the use of negative control exposures is that any tendency for reduced or exaggerated recall of exposure is likely to be similar for the main study exposure and the negative control exposure.

Although this approach can also be used to assess other types of bias (e.g. confounding), the focus in this section is on recall bias in case–control studies, as in Example 4.11.

Example 4.10. Using DAGs to identify recall bias

In the Interphone case–control study of mobile phone use and brain tumours (Cardis et al., 2007), researchers conducted a validation study on a subsample of the participants by comparing the self-reported mobile phone use with data from network operators (Vrijheid et al., 2009). The number of calls was underestimated, but the underestimation was similar among case and control participants, suggesting that there was non-differential misclassification for this exposure variable. In a DAG, this would translate into a lack of an arrow from the case status to the reported mobile phone use, as shown in Fig. 4.1, even if there were still factors that affected the reported exposure status other than the actual exposure. (text continues above)

Fig. 4.1. Directed acyclic graph of a study with underreporting of the prevalence of mobile phone use (exposure) but non-differential misclassification by brain tumour (outcome) status.



Example 4.11. Negative control exposures to assess recall bias in a study of pesticide exposure

In a case–control study of a particular pesticide (pesticide A) and cancer, any influences on the reporting of exposures (e.g. case participants being more likely than control participants to recall pesticide exposures) are likely to apply to pesticides in general, rather than only to pesticide A. If it is well established that another pesticide (pesticide B) is not associated with the cancer under study (e.g. if there had been a cohort study of workers predominantly exposed to this other pesticide), then pesticide B could serve as a negative control exposure. Thus, if a strong association was found between pesticide B and the outcome in the case–control study, this would provide evidence of information bias, as well as its likely magnitude and direction. (text continues below)

A related approach is the examination of positive control outcomes to assess the validity and quality of the exposure metric for an agent that has been found to be associated with other outcomes besides the one being investigated (Example 4.12).

(c) Tool E-3: examination of exposure information from different sources

In some instances, exposure information (e.g. from questionnaires) can be combined with more objective exposure measures. For example, determining whether participants have worked as a farmer would be a relatively poor measure of exposure to pesticides, but this can be ascertained reasonably accurately, through either questionnaires or examination of work history records. If, for example, there were recall bias with regard to exposure to pesticides, with case participants more likely than control participants to recall and report past exposures, one might expect this to be apparent in artificially high odds ratios when using exposure questionnaires, but one would not expect this bias to occur when "whether the participant has ever worked as a farmer" was the exposure metric; in this situation, taking the participant's being a farmer as the exposure might be expected to involve some non-differential information bias (which would usually be towards the null because the exposure is dichotomous) but would probably avoid or minimize differential recall bias. Similar considerations would apply when examining analyses restricted to exposures involving major events (e.g. work as a pesticide sprayer) rather than minor events (e.g. spray drift from a neighbouring farm).

(d) Tool E-4: comparisons with external data

Another approach for assessing information bias involves comparing the study data with external data on the prevalence of the exposure in the source population (Examples 4.13 and 4.14). This can involve information either on the exposure itself (e.g. smoking rates in the general population) or on a surrogate of the exposure. For example, if the exposure under study is the use of a pharmaceutical drug (prescribed or non-prescribed) and it is believed that control participants (but not case participants) may be underreporting, or not recalling, previous exposures, then one might compare the exposure prevalence in the control participants with that expected on the basis of general population rates of use.

(e) Tool E-5: consideration of analysis stratified by index versus proxy interviews

In studies involving proxy interviews, sensitivity analyses stratified on index interviews versus proxy interviews (i.e. interviews with the relevant case or control participant versus interviews with a proxy) can provide indirect evidence about whether the use of proxy interviews introduced

Example 4.12. Positive control outcomes to assess exposure misclassification in a study of benzene exposure

In an evaluation of whether benzene is a cause of lung cancer, *IARC Monographs* reviewers considered whether a cohort study demonstrated the expected association between benzene and leukaemia. A finding that the benzene exposure metric did not show this anticipated association for leukaemia led to scepticism of the adequacy of the exposure assessment (IARC, 2018). (text continues above)

Example 4.13. Using national statistics to assess recall bias

The European Union (EU) Labour Force Survey (Eurostat, 2022) reports statistics for the number of people working at night as a percentage of the total number of employed people in Europe, stratified by geopolitical entity, sex, age class, and calendar year. Similar data are available in other areas of the world. This information can be compared with the prevalences obtained for control participants in case–control studies on night shift work and cancer risk. Note that this is a rough comparison, because data would not be specific for the exact age distribution, study area, or study period. Nevertheless, these statistics can be used to identify the presence of major information bias problems. However, it should also be recognized that if such problems exist, they could reflect either information bias or selection bias (see Chapter 5). (text continues on page 98)

Example 4.14. Recruiting different types of control groups to assess recall bias

In *IARC Monographs* Volume 126, on opium use (IARC, 2021), the Working Group evaluated two case–control studies of oesophageal cancer (carried out by a single research team), in which different control groups were recruited: one hospital-based and one neighbourhood-based (Shakeri et al., 2012). The Working Group concluded that the neighbourhood-based control group probably provided a less biased estimate, because the prevalence of opium use reported by the neighbourhood-based control participants was similar to that reported from other sources for the general population of the region. (text continues on page 98)

information bias; however, such analyses entail strong assumptions. Typically, investigators report the full results and the results of the analysis restricted to the interviews with the index participants (because proxy interviews are used mainly or exclusively with case participants). If data from index participants are perfect (i.e. no exposure measurement error) or very nearly so, then conducting stratified analyses and estimating the exposure–outcome association among the index case participants can reduce bias. As shown by Greenland and Robins (1985), this approach has very important limitations. First, if the sensitivity and specificity are not perfect among the index case participants, there is no guarantee that this approach will yield less bias than an analysis that ignores the distinction between index and proxy responses. Second, such stratified analyses can increase the variance of study estimates; researchers need to weigh the benefits of a reduction in bias against a corresponding increase in variance. If such analyses are to be undertaken, it would be good practice to estimate the magnitude of bias under plausible sensitivity and specificity parameters for proxy and index case participants, as exemplified in Greenland and Robins (1985).

(f) Tool E-6: triangulation using comparisons across studies

Information bias from differential errors in exposure can also be assessed using triangulation approaches, introduced in Chapter 3, by making comparisons across studies. This applies particularly when similar studies have been conducted in the same population (e.g. cohort studies involving the same industry or the same group of workers, or case–control studies conducted in the same populations). However, comparisons can also be made between studies in different populations where it is reasonable to assume that the strength of the main exposure–outcome association is likely to be similar. For example, one might compare the findings from studies in which interviews were used to obtain exposure information with those from studies in which more objective methods, such as the analysis of personnel records on work history (e.g. Example 4.15), were used. Such comparisons across studies are discussed in Chapter 6.

Example 4.15. Using triangulation to assess recall bias

Two exposure assessment approaches were used in population-based case–control studies included in *IARC Monographs* Volume 124, on night shift work (IARC, 2020). The first approach typically used subjective methods (questionnaires and interviews) to assess the exposure to night shift work, to ascertain precise information on jobs held, as well as start and end times for each job (e.g. Papantoniou et al., 2016). The second approach used general population-based JEMs exclusively when characterizing exposure (e.g. Hansen, 2001). The Working Group considered the second approach to be prone to a large degree of exposure misclassification in assessing night shift work, because it would provide a highly imprecise measure of the exposure (i.e. with non-differential information bias, usually towards the null). Therefore, they excluded such studies from further consideration. In contrast, the second approach would avoid or minimize differential recall bias. Questionnaires provide more precise assessments of the individual exposure, but the reporting might be affected by knowledge of the outcome status, resulting in (differential) recall bias (most probably away from the null). The Working Group could have compared the findings of studies using these two methods to assess their respective possible biases (which might be expected to operate in different directions). (text continues on page 99)

4.3 Tools for quantifying bias due to errors in exposure

4.3.1 Tool E-7: simple bias analysis for exposure misclassification

Bias analyses of exposure misclassification for a binary (i.e. yes or no) exposure can be performed if one has information on the sensitivity and specificity of the exposure measurement method. These data may be available from an internal validation study or from external sources, such as previous validation studies published in the literature. Alternatively, expert opinion can be used to inform sensitivity and specificity parameters (Goldsmith et al., 2023). However, the quality of the bias analysis will be determined by the quality of the sensitivity and specificity parameters, so these assumptions should not be made lightly.

The formulae in Table 4.1 enable us to predict which data would be observed if the counts of correctly classified data and the accompanying sensitivities and specificities were known. In practice, only the observed cell counts are known, with perhaps estimates of sensitivities and specificities. Solving the four equations in Table 4.1 for the correctly classified cell counts results in the following simple formulae:

$$A = \frac{a - N_1(1 - \mathrm{sp}_1)}{\mathrm{se}_1 + \mathrm{sp}_1 - 1} \quad (4.1)$$

$$B = N_1 - A \quad (4.2)$$

$$C = \frac{c - N_0(1 - \mathrm{sp}_0)}{\mathrm{se}_0 + \mathrm{sp}_0 - 1} \quad (4.3)$$

$$D = N_0 - C \quad (4.4)$$

These formulae enable prediction of the data that would have been seen (correctly classified) given the observed cell counts and posited sensitivities and specificities.

This methodology is used in a spreadsheet for exposure misclassification (Chapter 6) that accompanies the textbook by Fox et al. (2021) (https://sites.google.com/site/biasanalysis/Home; the spreadsheet is provided in Annex 2, online only, available from: https://publications.iarc.who.int/634#supmat), as demonstrated in Examples 4.16 and 4.17.

4.3.2 Tool E-8: multidimensional analysis

A multidimensional sensitivity analysis can also be performed, in which various combinations of specificities or sensitivities in case and control participants are used to develop a range of bias-adjusted estimates (Fox et al., 2005; Johnson et al., 2014; Fox et al., 2023; Example 4.18).

Table 4.1. Relation between correctly classified (uppercase) and observed (lowercase) data in a case–control study with misclassification of exposure

	Correctly classified		Total	Observed data	
	Exposed	Unexposed		Exposed	Unexposed
Case participants	A	B	N_1	$a = se_1A + (1 - sp_1)B$	$b = (1 - se_1)A + sp_1B$
Control participants	C	D	N_0	$c = se_0C + (1 - sp_0)D$	$d = (1 - se_0)C + sp_0D$

se_0, sensitivity for control participants; se_1, sensitivity for case participants; sp_0, specificity for control participants; sp_1, specificity for case participants.

Example 4.16. Analysis of bias from non-differential exposure misclassification

Fritschi et al. (2013) conducted a population-based case–control study in Western Australia that examined the association between shift work and breast cancer risk. The study involved 1202 case participants who had incident breast cancer and 1785 frequency age-matched control participants who were identified between 2009 and 2011. A self-administered questionnaire was used to collect information on demographic, reproductive, and lifestyle factors and lifetime occupational history, and a telephone interview was used to obtain further details about shift work and lifestyle risk factors. Weak evidence of an increase in the risk of breast cancer was observed among women who worked night shifts (OR, 1.16; 95% CI, 0.97–1.39).

The investigators did not report estimates of the sensitivity or specificity of their exposure measure, but it is likely that there was some degree of misclassification, given that the exposures were based on questionnaire data. For this exercise, it is assumed that some individuals failed to understand the questions or may not have correctly answered the questions for other reasons. It is also assumed that these errors were non-differential with respect to disease.

A simple bias analysis can be performed using the methodology described in this section, assuming that the misclassification errors in the study were non-differential with respect to the disease and that there was a modest amount of error (sensitivity, 80%; specificity, 90%). The crude (i.e. unadjusted for measurement errors) results from the study and the results adjusted for misclassification bias are presented in Table 4.2. The crude (i.e. unadjusted) odds ratio is 1.16 (95% CI, 0.98–1.38), which is almost identical to the results adjusted for measured confounders (OR, 1.16; 95% CI, 0.97–1.39) presented in the paper. However, the odds ratio derived from the bias-adjusted data (OR, 1.29) was somewhat greater than the results without adjustment for misclassification, suggesting that misclassification of exposure may have biased the results towards the null. Confidence intervals for the misclassification-adjusted estimate are available from either Greenland (1988) or Chu et al. (2006).

$$\mathrm{Var}(\ln \mathrm{OR}) = \frac{N_1 ab(\mathrm{se}_1 + \mathrm{sp}_1 - 1)^2}{(N_1\mathrm{se}_1 - a)^2(N_1\mathrm{sp}_1 - b)^2} + \frac{N_0 cd(\mathrm{se}_0 + \mathrm{sp}_0 - 1)^2}{(N_0\mathrm{se}_0 - c)^2(N_0\mathrm{sp}_0 - d)^2} \quad \text{(E4.4)}$$

Table 4.2. Observed and misclassification-adjusted results from the case–control study of breast cancer by Fritschi et al. (2013) assuming non-differential errors and 80% sensitivity and 90% specificity

	Observed data		Total	Data adjusted for misclassification	
	Exposed	Unexposed		Exposed	Unexposed
Case participants	a = 288	b = 914	N_1 = 1202	A = 239.7	B = 962.3
Control participants	c = 381	d = 1404	N_0 = 1785	C = 289.3	D = 1495.7

In this problem, the resulting variance is 0.023, yielding a 95% confidence interval of (0.96, 1.73). This interval is slightly wider than the original interval; this is generally the result for bias analyses. (text continues on page 100)

Example 4.17. Analysis of bias from differential exposure misclassification

The same methodology as in Example 4.16 can be used to assess exposure misclassification that is differential with respect to disease. For example, Mohebbi et al. (2021) reported findings from a case–control study of head and neck squamous cell carcinoma (HNSCC) and opium use. The study included 633 case participants with head and neck cancer, who had been identified in cancer hospitals in 10 provinces in the Islamic Republic of Iran. Control participants (n = 3065) were hospital visitors, frequency-matched to the case participants on age, sex, and location. Mohebbi et al. (2021) assessed opium use with a standardized self-reported questionnaire. Overall, they reported an increased risk of HNSCC among regular opium users compared with non-users, with an adjusted odds ratio of 3.76 (95% CI, 2.96–4.79). Mohebbi et al. (2021) expressed concern over possible misclassification of opium use and performed preliminary sensitivity analyses in their study.

In a separate publication, Rashidian et al. (2017) conducted a cross-sectional hospital- and community-based validation study of self-reported opioid use, using a urine rapid screening test for opioid metabolites as a validation measure, in hospitals that were referral centres for cancer in 4 of the 10 provinces in the Islamic Republic of Iran that were included in the case–control study conducted by Mohebbi et al. (2021). This study involved patients who were hospitalized with chronic or acute conditions not related to opioid use, who were believed to have a similar referral pattern to the case participants, and healthy participants, who were selected from people accompanying patients with a chronic condition to a hospital in a manner similar to the method of selecting control participants used by Mohebbi et al. (2021). Rashidian et al. (2017, Figure 1) reported results that yielded a sensitivity of 79% and a specificity of 83% among hospitalized patients and a sensitivity of 68% and a specificity of 93% among healthy participants for self-reported opioid use compared with urine analysis. Note that Rashidian et al. (2017) used a composite outcome (urine analysis and thin-layer chromatography) as their gold standard, but in this example only urine analysis is used, for ease of presentation.

An adjustment for bias due to the differential misclassification of exposures in the study of Mohebbi et al. (2021) can be performed using the estimates of sensitivity and specificity given by Rashidian et al. (2017) and the statistical methodology described in this section and in Fox et al. (2021). The crude (i.e. unadjusted for either confounding or misclassification) results from the study and the results adjusted for misclassification bias are presented in Table 4.3. The crude (i.e. unadjusted) odds ratio from this study is 5.33 (95% CI, 4.42–6.41), and the misclassification-bias-adjusted odds ratio is 7.19 (95% CI, 5.17–10.00). It is noteworthy that both the crude and misclassification-adjusted results are substantially greater than the confounding-adjusted results presented by Mohebbi et al. (2021) (OR, 3.76; 95% CI, 2.96–4.79). This suggests that the confounding-adjusted results are biased towards the null due to exposure misclassification, and also that the crude and misclassification-adjusted results appear to be biased by confounding, because the crude result differs from the confounding-adjusted result. (text continues on page 100)

Table 4.3. Observed and misclassification-adjusted crude results from Mohebbi et al. (2021) using estimates of sensitivity and specificity from Rashidian et al. (2017)

	Observed data		Total	Data adjusted for misclassification	
	Exposed	Unexposed		Exposed	Unexposed
Case participants	a = 295	b = 368	N_1 = 663	A = 294.0	B = 369.0
Control participants	c = 401	d = 2664	N_0 = 3065	C = 305.7	D = 2759.3

Example 4.18. Multidimensional sensitivity analysis

In the validation study by Rashidian et al. (2017), 45 of 57 hospitalized people whose urine tested positive for opioids also reported use of opioids. From this, we can calculate a sensitivity of 79% with a 95% confidence interval of 66–89%. Repeating this for specificity, we obtain a specificity of 83% and a 95% confidence interval of 76–90%. Among healthy individuals in the validation study, we obtain a sensitivity of 68% (95% CI, 50–82%) and a specificity of 93% (95% CI, 87–96%). The sensitivity of the misclassification-adjusted odds ratio from Mohebbi et al. (2021) to the chosen values of sensitivity and specificity can be investigated by repeating this bias analysis using the estimated upper and lower confidence bounds of sensitivity and specificity. These values were chosen because they represent the limits of the sensitivity and specificity values supported by the validation data and therefore the most "extreme" possibilities. The results from the multidimensional analysis are shown in Table 4.4. At the lower limit of specificity among the control participants (87%), almost all control participants who reported opioid use are assumed to have been misclassified, and the misclassification-adjusted number of exposed control participants is quite small, resulting in implausibly large misclassification-adjusted odds ratios. The remaining permutations of the bias parameters all result in elevated odds ratios; however, four sets of values result in adjusted odds ratios that are nearer to 1 than the crude estimate. This illustrates how with differential misclassification one can have results that are biased either towards or away from the null. (text continues on page 104)

Table 4.4. Multidimensional analysis of data on opioid use and head and neck squamous cell carcinoma from Mohebbi et al. (2021), adjusted for misclassification of self-reported opioid use

Bias parameter				Adjusted cell count				OR_{adj}
se_1	sp_1	se_0	sp_0	*A*	*B*	*C*	*D*	
1	1	1	1	295.0	368.0	401.0	2664.0	5.33
0.66	0.76	0.5	0.87	323.5	339.5	6.9	3058.1	422.87
0.89	0.76	0.5	0.87	209.0	454.0	6.9	3058.1	204.34
0.66	0.9	0.5	0.87	408.4	254.6	6.9	3058.1	711.74
0.89	0.9	0.5	0.87	289.5	373.5	6.9	3058.1	343.92
0.66	0.76	0.82	0.87	323.5	339.5	3.7	3061.3	789.43
0.89	0.76	0.82	0.87	209.0	454.0	3.7	3061.3	381.46
0.66	0.9	0.82	0.87	408.4	254.6	3.7	3061.3	1328.69
0.89	0.9	0.82	0.87	289.5	373.5	3.7	3061.3	642.03
0.66	0.76	0.5	0.96	323.5	339.5	605.2	2459.8	3.87
0.89	0.76	0.5	0.96	209.0	454.0	605.2	2459.8	1.87
0.66	0.9	0.5	0.96	408.4	254.6	605.2	2459.8	6.52
0.89	0.9	0.5	0.96	289.5	373.5	605.2	2459.8	3.15
0.66	0.76	0.82	0.96	323.5	339.5	356.9	2708.1	7.23
0.89	0.76	0.82	0.96	209.0	454.0	356.9	2708.1	3.49
0.66	0.9	0.82	0.96	408.4	254.6	356.9	2708.1	12.17
0.89	0.9	0.82	0.96	289.5	373.5	356.9	2708.1	5.88

OR_{adj}, adjusted odds ratio; se_0, sensitivity for control participants; se_1, sensitivity for case participants; sp_0, specificity for control participants; sp_1, specificity for case participants.

4.3.3 Limitations of methods for analyses of exposure measurement errors

A major limitation of these methods that were used to conduct sensitivity analyses or adjust for misclassification errors is that they all involve using the crude results (i.e. unadjusted results) from the studies and thus ignore potential bias due to confounding. This is not problematic when the crude results are nearly equivalent to the results from the adjusted analyses, as seen in the study by Fritschi et al. (2018). However, Mohebbi et al. (2021) found evidence of confounding: the crude odds ratio (5.33; 95% CI, 4.42–6.41) and the confounding-adjusted odds ratio (3.76; 95% CI, 2.96–4.79) are appreciably different. A technically appropriate adjustment for confounding and exposure misclassification requires access to individual-level data. Such approaches are explained in detail in Fox et al. (2021). In practice, an *IARC Monographs* Working Group may be interested in adjusting for confounding (see Chapter 3) and misclassification but will generally only have access to aggregate data. In this situation, an approximate approach that can be used to adjust for confounding is to compute the ratio of the adjusted and crude odds ratios, ignoring misclassification, and apply that ratio to the misclassification-adjusted odds ratio, as demonstrated in Example 4.19. See Chapter 6 for further discussion of multiple-bias analysis.

4.3.4 Tool E-9: multiple categorical bias analysis

A similar approach to that used for binary exposures (Sections 4.3.1 and 4.3.2) could be taken for a study with a larger number of categories of exposure. To do this, one would have to know the percentage of individuals who were incorrectly classified in each category, and into which category they were inappropriately classified. This type of information is less likely to be available in epidemiological publications and would be particularly difficult to obtain for studies with a large number of categories, or where categories are unique to a particular study. However, assuming that the information is available, one could use this method to conduct a sensitivity analysis (Example 4.20).

The results from this sensitivity analysis do not suggest a monotonic decrease in risk with increasing duration of exposure, as was observed in the results reported in the study.

4.3.5 Tool E-10: probabilistic bias analysis

As mentioned in Sections 4.3.1 and 4.3.2, one or more values of the bias parameters must be specified when quantifying bias. The approach described in this section, probabilistic bias analysis, is an extension of multidimensional bias analysis and enables incorporation of the uncertainty in the bias parameters into the measures of association. In practice, probabilistic bias analysis involves specifying a probability distribution for each bias parameter that represents the uncertainty in the values. Samples are repeatedly drawn from each bias parameter distribution, and a simple bias analysis is repeated for each set of sampled bias parameters.

Example 4.19. Sensitivity analysis for both confounding and misclassification

For the study by Mohebbi et al. (2021), the ratio of the confounding-adjusted odds ratio to the crude odds ratio is 3.76/5.33 = 0.705. This ratio is the extent to which the observed crude odds ratio is altered after adjusting for confounding, and it can be applied to the misclassification-adjusted odds ratios calculated previously. For example, when adjusting for misclassification of opioid use, a misclassification-adjusted odds ratio of 7.19 was found. Multiplying this effect by the ratio of the confounding-adjusted odds ratio to the crude odds ratio gives an approximate estimate of a confounding- and misclassification-adjusted odds ratio of 7.19 × 0.705 = 5.07. Adjustment for misclassification bias increased the odds ratio, whereas adjustment for confounding bias decreased the odds ratio. In this example, the two sources of bias nearly cancel each other out, resulting in a bias-adjusted odds ratio that is very similar to the crude odds ratio. However, this will not always be the situation. (text continues above)

Example 4.20. Sensitivity analysis for categorical exposure misclassification

Fritschi et al. (2013) conducted a population-based case–control study that examined the association between shift work and breast cancer risk (as described in Section 4.3.1). An inverse exposure–response relation was observed in the study for duration of work in the night shift and breast cancer risk, as summarized in Table 4.5.

Table 4.5. Association between duration of exposure to working in the night shift and breast cancer risk (Fritschi et al., 2013)[a]

Duration of exposure	Case participants	Control participants	Crude OR (95% CI)	Age-adjusted OR (95% CI)
Never	914	1404	Reference	Reference
< 10 years	164	199	1.27 (1.01–1.59)	1.25 (1.00–1.56)
10 to < 20 years	71	98	1.11 (0.80–1.54)	1.09 (0.79–1.50)
≥ 20 years	53	84	0.97 (0.67–1.40)	1.02 (0.71–1.45)

CI, confidence interval; OR, odds ratio.

[a] Crude odds ratios were estimated using data presented in Table 2 in Fritschi et al. (2013). Confidence intervals were estimated using exact methods.

To check whether exposures were being underestimated in this study, a sensitivity analysis might be conducted, with the assumption that 20% of each category belonged in the next highest category. This would yield the adjusted results presented in Table 4.6. (text continues on page 104)

Table 4.6. Sensitivity analysis, assuming that 20% of case and control participants in each category should be in the next highest exposure group

Duration of exposure	Case participants	Control participants	Misclassification-adjusted odds ratio
Never	731.2	1123.2	Reference
< 10 years	314.0	440.0	1.10
10 to < 20 years	89.6	118.2	1.16
≥ 20 years	67.2	103.6	1.00

The uncertainty in the bias parameters is thus taken into account in the resulting error-adjusted estimates. The distribution of the error-adjusted estimates gives the analyst a more complete idea of the distribution of plausible effects than can be obtained through simple bias analysis or multidimensional bias analysis, and it is used to derive point and interval estimates, such as the median or the 95% simulation interval (i.e. the interval between the 2.5th and the 97.5th percentiles). Probabilistic bias analysis relies on the assumption that the specified bias parameter distributions are valid. Fox et al. (2021) provide more detailed information about probabilistic bias analysis and extend the idea of probabilistic bias analysis outlined here by incorporating random error introduced by the data collection process in addition to systematic error arising from misclassification (the accompanying spreadsheets as well as SAS and R code help facilitate application of the method; see Fox et al., 2021 and https://sites.google.com/site/biasanalysis/Home; R code is provided online only, available from: https://publications.iarc.who.int/634#supmat); see Example 4.21.

4.3.6 Tool E-11: regression calibration for continuous and categorized measures of exposure

(a) Continuous measures of exposure

In Section 4.2.1 it was discussed how errors in exposure measurement might cause bias in the estimated associations of the exposure with health outcomes. Regression calibration (Rosner et al., 1990; Section 5 of Keogh et al., 2020) is a statistical method to account for non-differential measurement errors in an exposure that is measured on a continuous scale, yielding an estimate that, in the best circumstances, is free from such bias, or at least has bias that is considerably reduced (Example 4.22a).

Regression calibration can be used to provide adjustment for non-differential measurement errors in epidemiological models. Simple regression calibration requires the following three basic steps. (To keep the description simple, confounder variables are not shown in the models.)

- Step (i). Regress the outcome (Y) on the measured exposure (X^*) to obtain a raw estimate of the association through a rate ratio or a hazard ratio. For example, the outcome model may be a Cox regression model, $h(t) = h_0(t)\exp(\beta_1 X^*)$, where $h(t)$ is hazard of an event ($Y = 1$) at time t and the association is measured as β_1, the log hazard

Example 4.21. Probabilistic bias analysis for exposure misclassification

The example described in Sections 4.3.1 and 4.3.2 on the association between differentially misclassified opium use and HNSCC provides a good illustration of probabilistic bias analysis. To express the uncertainty in each of the bias parameters, a triangular distribution is used as the bias parameter distribution, with the most probable values from Rashidian et al. (2017) as the mode and the respective limits of the 95% confidence intervals as the limits of the triangular distribution (Table 4.7). Probabilistic bias analysis is applied, as described in Fox et al. (2021), to account for random and systematic errors. We assumed no correlation between sensitivities among cases and controls or between specificities among cases and controls, although other assumptions are available.

Table 4.7. Parameters of triangular distributions used as bias parameter distributions for probabilistic bias analysis of data from Mohebbi et al. (2021) on misclassified opium use and head and neck squamous cell carcinoma

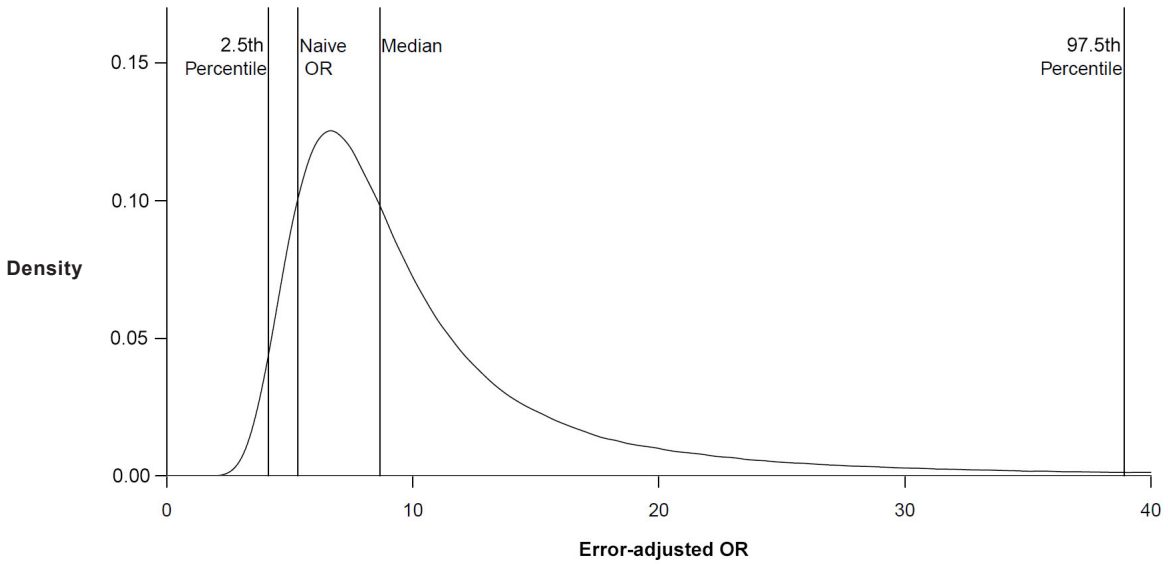
Bias parameter	Distribution parameters of triangular distribution		
	Minimum (%)	Mode (%)	Maximum (%)
se_1	66	79	89
sp_1	75	83	90
se_0	50	68	82
sp_0	87	93	96

se_0, sensitivity for control participants; se_1, sensitivity for case participants; sp_0, specificity for control participants; sp_1, specificity for case participants.

Fig. 4.2 shows the distribution of the error-adjusted odds ratios from 100 000 iterations. The median error-adjusted odds ratio is 8.66, with a 95% simulation interval of 4.13–38.9. About 89% of the error-adjusted odds ratios are greater than the unadjusted odds ratio of 5.33, indicating a bias towards the null in the analysis of Mohebbi et al. (2021). Because the 95% simulation interval is much wider than the 95% confidence interval of the unadjusted odds ratio, and because neglecting random error changes the error-adjusted odds ratio only slightly (median error-adjusted OR, 8.62; 95% simulation interval, 4.41–37.33, based on 10 000 iterations), the potential effect of systematic error due to exposure misclassification on the analysis is stronger than the effect of random error. This bias analysis offers some confirmation that the positive association in Mohebbi et al. (2021) is not a spurious finding from exposure misclassification, and it also highlights the extreme uncertainty around the magnitude of effect after adjusting for misclassification. (text continues above)

Example 4.21. Probabilistic bias analysis for exposure misclassification (continued)

Fig. 4.2. Distribution of error-adjusted odds ratios (ORs) resulting from probabilistic bias analysis of data from Mohebbi et al. (2021) on misclassified opium use and head and neck squamous cell carcinoma.



ratio for a unit increase in the measured exposure (Example 4.22b).

- Step (ii). An attenuation factor, usually denoted by λ, is estimated from some validation data. The simplest way to estimate λ is to obtain a reference (gold standard) measure of the exposure (X) in a subgroup of participants and perform a linear regression of X on X^*: $X = \lambda_0 + \lambda X^* + \varepsilon$. This model is called the calibration model, and the attenuation factor is estimated as the regression coefficient, λ, of X^* (λ_0 represents an offset value, and ε represents the error term). When reference measurements are not available, even in a subgroup of participants, the attenuation factor might be estimated from data that are external to the study (Example 4.22c). When external data are used to estimate the attenuation factor, the study being analysed and the external study must be similar with respect to the main assessment instrument used to measure the exposure, the distribution of exposure among the population, and the covariates used for adjustment.
- Step (iii). The association is adjusted for measurement error by dividing the estimated association parameter β_1 by the estimated attenuation factor; in mathematical notation, $\beta_{1\text{-adjusted}} = \beta_1/\lambda$ (Example 4.22d).

These three steps form the core of the regression calibration method in its simplest form. Different types of validation data can be used when

Key message

In most applications, as in Example 4.22, the attenuation factor (λ) in regression calibration is positive and less than 1, and usually ranges between 0.3 and 0.7, indicating, respectively, limited and adequate accuracy of the observed assessments compared with the truth. Therefore, the adjustment of dividing by λ inflates, or de-attenuates, the estimated association. Sensitivity analyses using a range of estimates for this attenuation factor (e.g. 0.3–0.7) can provide an understanding of the magnitude of the underestimation of the risk due to measurement error.

Example 4.22a. Regression calibration for adjustment for measurement error

Within the Swedish Mammography Cohort, a rate ratio for colorectal cancer incidence of 1.20 (95% CI, 0.99–1.45) was reported for an increase of 100 g/day of red meat intake (Larsson et al., 2005). Red meat intake was based on dietary intake, self-reported in a food frequency questionnaire, which was subject to measurement errors. The estimated rate ratio needed to be adjusted for these errors. (text continues on page 106)

Example 4.22b. Regression calibration for adjustment for measurement error (continued)

In the Swedish Mammography Cohort, β_1 was estimated as ln(1.20) = 0.18. (text continues on page 107)

Example 4.22c. Regression calibration for adjustment for measurement error (continued)

In the Swedish Mammography Cohort, no reference measurements were available. However, an attenuation factor could be estimated from data collected within the European Prospective Investigation into Cancer and Nutrition (EPIC), a large prospective study with more than 500 000 participants recruited in 10 European countries (Riboli et al., 2002). Reference measurements based on 24-hour recall data obtained from a subset of 36 994 participants were used to estimate an attenuation factor for food frequency questionnaire self-reported red meat intake of 0.51. (text continues on page 107)

Example 4.22d. Regression calibration for adjustment for measurement error (continued)

In the Swedish Mammography Cohort, the adjusted log hazard ratio was estimated as ln(1.20)/0.51 = 0.357; from this value, the adjusted hazard ratio may be estimated as exp(0.357) = 1.43. (text continues on page 107)

Example 4.22e. Estimating an adjusted confidence interval with regression calibration

In the Swedish Mammography Cohort, the unadjusted hazard ratio for colorectal cancer per increment of 100 g/day of red meat intake was reported as 1.20, with a 95% confidence interval of 0.99–1.45. Thus, the confidence interval for the log hazard ratio of 0.18 was ln(0.99) to ln(1.45), that is, from −0.01 to 0.37. The attenuation factor, λ, that was used for adjustment was 0.51. A simple approximate way of estimating the confidence limits for the adjusted log hazard ratio is to divide by λ, giving −0.02 to 0.73. Converting back to the hazard ratio scale, by exponentiating, gives a 95% confidence interval of 0.98–2.07 for the adjusted hazard ratio (recall that its value was 1.43). (text continues on page 109)

estimating the attenuation factor, depending on the type of measurement error (see Section 4 of Keogh et al., 2020). This description does not include other covariates in the exposure–outcome model or in the exposure calibration model. Any other covariates that are included in the outcome model should also be included in the calibration model. In Example 4.22, and for most external validation data, the attenuation factor is derived from a calibration model that does not include the same covariates as the outcome model. In that situation, the estimated attenuation factor must be regarded as an approximation that may carry some bias.

Within the context of expert reviews, such as *IARC Monographs* evaluations, an important constraint is that the implementation of regression calibration must usually rely on external data, because attenuation factors are not reported for most studies. Therefore, the resulting adjusted estimate of the association parameter should be regarded as a ballpark estimate. For an example of regression calibration carried out using the original study data, as recommended wherever possible, see the description of a study of red meat consumption and colorectal cancer in Section 7.4.3.

Approximate upper and lower confidence limits for the adjusted association can also be estimated. In mathematical notation, if L_1 and L_2 are the upper and lower confidence limits for the association parameter β_1 (in Example 4.22, the log hazard ratio), then the adjusted confidence limits are L_1/λ and L_2/λ (Example 4.22e).

As shown in Example 4.22e, the regression calibration adjustment makes the confidence interval wider, expressing the extra uncertainty in the estimated association caused by the measurement error. Note also that, using this method, if the unadjusted confidence interval for the association covers the null value, the adjusted confidence interval will still cover the null value. Thus, in general, this ballpark adjustment will not alter the judgement of whether the association is statistically significant, but, importantly, it will provide a better understanding of the likely magnitude of the association.

Note that this method of adjusting the confidence interval for the association is approximate and does not take into account the uncertainty in the estimate of the attenuation factor, λ. Rosner et al. (1989) give a method of incorporating this uncertainty into the confidence interval, which makes the interval still wider than the one estimated from the simple method provided here. For expert reviews in which access to original study data is lacking, the method of Rosner et al. (1989) could be used, but only when the attenuation factor estimate that is available is accompanied by an estimate of its standard error. In mathematical notation, suppose that the standard error of λ is s and the standard error of the unadjusted estimate of the association parameter β_1 is se, and that its 95% confidence limits, as before, are denoted by L_1 and L_2. Then the lower confidence interval of the adjusted association parameter is given by

$$\frac{L_1}{\lambda} - \left(\frac{1.96}{\lambda}\right)\left(\sqrt{se^2 + \frac{\beta_1^2 s^2}{\ddot{e}^2}} - se\right) \quad (4.5)$$

and the upper confidence interval is given by

$$\frac{L_2}{\lambda} + \left(\frac{1.96}{\lambda}\right)\left(\sqrt{se^2 + \frac{\beta_1^2 s^2}{\ddot{e}^2}} - se\right) \quad (4.6)$$

When s, the standard error of λ, is set to zero, the formulae revert to the adjusted limits L_1/λ and L_2/λ given by the simpler method described previously.

To conclude this subsection, note that caution must be taken in using attenuation coefficients from substudies that use a self-report instrument, albeit one that is more accurate than the main study self-report instrument, as a reference measure. In the example of the EPIC study given here, 24-hour recall data were used as a reference measure for a food frequency questionnaire. The errors on two self-report instruments will often be correlated, introducing bias in the estimate of the attenuation coefficient. However, in dietary studies there is usually no feasible alternative, except for a limited number of nutrients, such as energy, protein, potassium, and sodium, for which reference biomarkers can be used.

(b) Categorized measure of exposure: mobile phone use and gliomas

The ballpark adjustment using the attenuation factor, as described in Section 4.3.6(a), is applicable when the exposure variable used in the exposure–outcome association model is continuous. However, the exposure–outcome association parameter is often expressed in terms of categorized exposure variables, for example when the continuous exposure is transformed into quintiles of its distribution. In nutritional epidemiology, it is quite common to report the relative

risk of a disease in the highest quintile of the dietary intake compared with the lowest quintile.

Key message

The approximate adjustment is achieved by using, in place of the attenuation coefficient, the correlation coefficient between the continuous true and observed exposures (Kipnis and Izmirlian, 2002), sometimes referred to as the validity coefficient. In other words, for categorized exposures, the association parameter estimated from the observed exposure can be adjusted for measurement error by dividing the estimate by the correlation coefficient, instead of by the attenuation factor.

Example 4.23 illustrates this type of adjustment.

4.3.7 Tool E-12: other methods for quantifying bias

In this section, three methods that are commonly used to adjust estimates for exposure measurement error – simulation extrapolation (SIMEX), the Bayesian method, and multiple imputation – are described in Side Boxes 4.2, 4.3, and 4.4, as other methods for quantifying bias due to exposure measurement error. However, because these approaches generally require individual-level data, they are only briefly outlined here with regard to summary-level data. Table 4.8 describes the process descriptions and situations in which these methods are preferable to those described previously.

4.4 Outcome misclassification

4.4.1 Non-differential outcome misclassification

In cancer epidemiology studies, outcome misclassification is not as common an issue as exposure misclassification but may still occur under some circumstances (Example 4.26).

Like mismeasurement of the exposure, misclassification or measurement error in the outcome can also bias results in epidemiological studies.

Example 4.23. Bias adjustment for misclassified categorical exposures

Momoli et al. (2017, Table 5) analysed the Canadian data of the 13-country case–control Interphone study (INTERPHONE Study Group, 2010), reporting an estimated odds ratio of 2.0 (95% CI, 1.2–3.4) for glioma among the category of participants reporting a lifetime cumulative mobile phone use of more than 558 hours, compared with a reference category (reporting never use, irregular use, use only within a year before the reference date, or use only with a hands-free device). The odds ratio estimate was derived from a conditional logistic regression model, adjusting for age, sex, region, education level, and interview lag. The simple ballpark adjustment of this odds ratio estimate for non-differential random error in exposure measurements is considered here.

Recall that the estimated association parameter is to be divided by the correlation coefficient between measured and true exposure. Vrijheid et al. (2006) describe a validation study in which data from 672 Interphone participants who reported cumulative hours of mobile phone use were compared with records obtained from their network operators, assumed to be their true exposure. The study-wide correlation coefficient between reported and true use measured on the logarithmic scale was 0.69, where recall was approximately 6 months after the actual use.

To perform the adjustment, first the odds ratio (2.0) and its confidence limits (1.2, 3.4) are converted to the natural log scale, because they are originally estimated from a logistic regression model:

ln OR = 0.69; 95% CI = (0.18, 1.22) (E4.5)

These values are then divided by the correlation coefficient, 0.69:

adjusted ln OR = 1.00; adjusted 95% CI = (0.26, 1.77) (E4.6)

Finally, these values are converted back to the original scale, by taking their exponent:

adjusted OR = 2.7; adjusted 95% CI = (1.3, 5.9) (E4.7)

Thus, after adjusting for non-differential random measurement error, the estimated odds ratio is increased from 2.0 to 2.7, and its confidence interval is considerably wider, especially at the upper end. (text continues above)

Table 4.8. Methods of adjustment for measurement error and situations in which they may be preferred

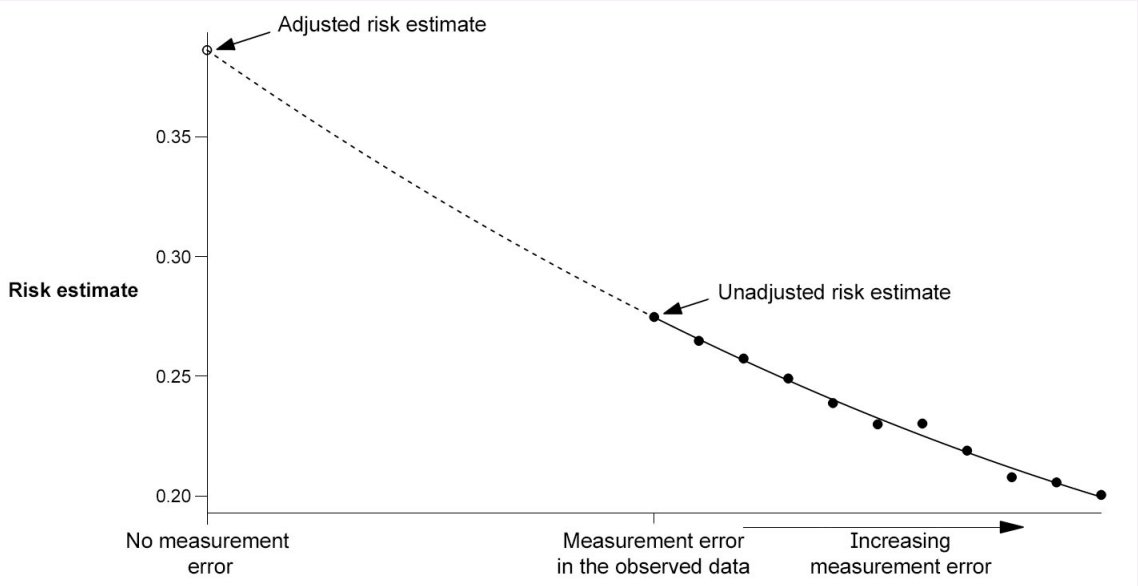
Method	Process description	Preferable in the following situations
Probabilistic bias analysis	Bias parameters are simulated.	Original study data are unavailable Bias model is known
MC-SIMEX	Increasing misclassification is simulated.	Exposure variable with more than two categories Multiple regression models
Bayesian method	Bias parameters, risk parameters, and other model parameters are simulated.	Integration of prior knowledge about model parameters other than bias parameters Flexible specification of the model beyond standard choices
Multiple imputation	The missing true exposure values are simulated.	Internal validation data are available Flexible specification of the risk model Bias model is unknown

MC-SIMEX, simulation extrapolation for misclassification.

Side Box 4.2. Simulation extrapolation for misclassification (MC-SIMEX)

In general, SIMEX (Cook and Stefanski, 1994) is a two-step approach: simulation and extrapolation. In the simulation step, the relation between the magnitude of the measurement error and the unadjusted risk estimate is approximated. For this purpose, the unadjusted regression model (e.g. a logistic regression model) is estimated several times using exposure data with gradually increasing measurement error. In the extrapolation step, the relation between the magnitude of the measurement error and the unadjusted risk estimates is extrapolated to the situation with no measurement error, yielding the error-adjusted risk estimate (see Fig. 4.3).

Fig. 4.3. Risk estimation using simulation extrapolation (SIMEX). Solid circles, unadjusted risk estimates based on observed and simulated data. Open circle, adjusted risk estimate. Solid line, model for the relation between the magnitude of the measurement error and the unadjusted risk estimates. Dashed line, extrapolation of the model to the situation with no measurement error.



Side Box 4.2. Simulation extrapolation for misclassification (MC-SIMEX) (continued)

The SIMEX for misclassification (MC-SIMEX) method is based on the SIMEX concept; the main differences are that the error-prone variable X^* is a discrete variable with k categories and that the magnitude of the measurement error is specified by the $k \times k$ misclassification matrix $\boldsymbol{\Pi}$ (Küchenhoff et al., 2006). In the situation of a single misclassified binary variable, the misclassification matrix can be determined using sensitivity and specificity:

$$\boldsymbol{\Pi} = \begin{pmatrix} \text{specificity} & 1 - \text{sensitivity} \\ 1 - \text{specificity} & \text{sensitivity} \end{pmatrix} \quad \text{(E4.8)}$$

The two steps in MC-SIMEX are simulation and extrapolation (Küchenhoff et al., 2006).

- Simulation: Simulate data with gradually increasing misclassification by reclassifying the observed data. Estimate the unadjusted regression model for each magnitude of misclassification.
- Extrapolation: Fit a parametric model for the unadjusted risk estimates depending on the magnitude of misclassification. Extrapolating this model to the situation with no misclassification yields the error-adjusted risk estimate.

Applications of this method can be found, for example, in Heid et al. (2008), Slate and Bandyopadhyay (2009), and Costas et al. (2015).

In contrast to the previously mentioned methods, MC-SIMEX can be used for an exposure variable with more than two categories and for multiple regression models. In addition, the approach to bias analysis with MC-SIMEX is very different from other bias analysis methods: all the necessary information about the misclassification is given in the misclassification matrix, so there is no need to specify a bias model. (text continues on page 110)

Key message

Bias from outcome misclassification is generally expected to be towards the null if the errors are non-differential with respect to exposure (i.e. there is no association between exposure and the misclassification errors).

It is worth emphasizing that, as with non-differential exposure misclassification, bias towards the null from non-differential outcome misclassification is only an expectation; the results from an individual study could be biased away from the null due to random error.

In epidemiological studies of cancer, outcome misclassification may arise for several reasons. In studies that rely on cancer or death certificate registries, misclassification can result from error-prone data in the registries related to changes in diagnostic codes, incomplete data, or data coding errors. For certain cancers, there may also be problems with imperfect sensitivity and specificity (Example 4.27).

Outcome misclassification can also result when tumour characteristics are overlooked, for example histological subtype or hormone receptor status (e.g. breast cancer) or aggressiveness (e.g. prostate cancer), which can have different risk factors, or from cancer misdiagnosis (e.g. peritoneal mesothelioma misdiagnosed as ovarian cancer), as in Example 4.28. This will be problematic if an exposure is exclusively or disproportionately associated with only one cancer subtype.

4.4.2 Differential outcome misclassification

Outcome classification errors that are differential with respect to exposure can bias results in either direction (Example 4.29).

4.4.3 Quantitative assessment of bias due to outcome misclassification

The methods described in Sections 4.3.1 and 4.3.2 can also be used to conduct sensitivity analyses of outcome misclassification based on assumptions about sensitivity and specificity or using data from a validation study (Gilbert et al., 2016). Analyses based on the cancer screening history of study subjects can also help to capture the magnitude of errors resulting from outcome misclassification (Example 4.30).

Side Box 4.3. Bayesian methods for error adjustment

A Bayesian approach allows for a very flexible consideration of the uncertainty regarding the bias parameters (e.g. dependencies between bias parameters). Bayesian approaches are used to estimate the distribution of the model parameters of interest from the prior distributions of the unobserved quantities and the data. A Bayesian model for quantifying bias consists of three model components (Fox et al., 2021):

- the risk model, i.e. the regression model, for the observed data;
- the bias model, i.e. the model describing the relation of the parameters in the risk model for the observed data and the corresponding error-free parameters; and
- the prior distributions for the unobserved quantities.

The prior distributions for the bias parameters included in the third model component correspond to the probability distributions for the bias parameters in the probabilistic bias analysis (Section 4.3.5). We chose truncated normal distributions for this example, but non-truncated normal distributions will generally be preferred. Application of both the Bayesian and probabilistic approaches requires a high degree of understanding and care (Fox et al., 2021).

The Bayesian model components for non-differential exposure misclassification in a case–control study are given in Example 4.24. The numbers of people observed to be exposed among case and control participants are modelled using binomial distributions, providing the odds ratio as a risk measure in the risk model. The relations between the proportions of the truly exposed and those observed to be exposed among case and control participants are described using sensitivity and specificity as bias parameters in the bias model. Because the error is non-differential, sensitivity and specificity do not differ between case and control participants. Independent beta distributions are chosen as prior distributions for the sensitivity and specificity.

This Bayesian model for quantifying bias includes both the parameters of the risk model, from which the carcinogenic risk estimate can be derived, and the bias parameters. In addition to prior information about the bias parameters, which is equivalent to the distribution placed on the sensitivity and specificity in probabilistic bias analysis, Bayesian methods can use prior distributions of other parameters (e.g. the risk parameter). Because Bayesian methods themselves already involve iterative sampling of data and parameters, their application for quantifying bias comprises only a single modelling step, which accounts simultaneously for the uncertainties in the parameters of the risk model and the bias parameters. More details on the difference between the Bayesian and probabilistic approaches to quantifying bias due to exposure misclassification can be found in Chu et al. (2006), MacLehose and Gustafson (2012), and Corbin et al. (2017). (text continues on page 110)

Example 4.24. Bayesian model components for non-differential exposure misclassification in a case–control study

Observed data

a	Number of people observed to be exposed among case participants	N_1	Number of case participants
c	Number of people observed to be exposed among control participants	N_0	Number of control participants

(1) Risk model	(2) Bias model	(3) Prior distributions
$a \sim \text{Binomial}(N_1, p_1^*)$	$p_1^* = p_1 se + (1 - sp)(1 - p_1)$	$p_1 \sim \text{Beta}(\alpha_1, \beta_1)$
$c \sim \text{Binomial}(N_0, p_0^*)$	$p_0^* = p_0 se + (1 - sp)(1 - p_0)$	$p_0 \sim \text{Beta}(\alpha_2, \beta_2)$
Unadjusted risk estimate:	Error-adjusted risk estimate:	$se \sim \text{Beta}(\alpha_3, \beta_3)$
$\text{OR}^* = \frac{\frac{p_1^*}{1-p_1^*}}{\frac{p_0^*}{1-p_0^*}}$	$\text{OR} = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}}$	$sp \sim \text{Beta}(\alpha_4, \beta_4)$

Source: Adapted from Fox et al. (2021).

Side Box 4.4. Multiple imputation

Exposure measurement errors can be considered to be a problem of missing data: true exposure values are missing. Therefore, methods of accounting for missing data, such as multiple imputation, can be used to calculate error-adjusted estimates directly and to quantify bias due to exposure measurement error (Greenland, 2009). A prerequisite for the use of multiple imputation is the availability of adequate prior information on the true exposure values, usually in the form of internal validation data for a subset of individuals. From this, an imputation model for the true exposure is estimated in conjunction with the other study data (e.g. outcome and observed exposure). Random draws are generated based on the imputation model and serve as true exposure values (imputation). These are then used to calculate a risk estimate (estimation). Imputation and estimation are repeated several times, and the error-adjusted risk estimate is obtained by combining the risk estimates from the individual iterations, as shown in Example 4.25.

Example 4.25. Opium use and HNSCC – bias analysis for categorical data

Quantifying bias due to misclassification using SIMEX, the Bayesian method, or multiple imputation usually requires the original study data. Only a very few scientific publications provide sufficient information for the application of these methods. To provide insight into the application of the Bayesian method and SIMEX, we again examine the example from Sections 4.3.1–4.3.3 and 4.3.5 on differentially misclassified opium use and HNSCC (Mohebbi et al., 2021). Multiple imputation cannot be used, because of a lack of internal validation data; as a way of working around this constraint, artificial validation data were generated and multiple imputation could then be applied to the example in this section, using the artificial validation data that had been generated.

The three components of Bayesian bias analysis are the same as in the example in Side Box 4.3. To apply this model, one must specify these components. The risk model results from the original scientific publication, and the bias model results from theoretical considerations. The prior distributions are selected during the bias analysis. Because there is no prior knowledge about the true proportions of exposed individuals among case participants (p_1) or among control participants (p_0), uninformative uniform priors with parameters 0 and 1 are chosen; this is equivalent to a beta distribution with both parameters equal to 1 ($\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1$). Truncated normal distributions are used as the prior distributions for the bias parameters, i.e. sensitivity and specificity among case and control participants. As in Section 4.3.5, the distribution parameters are derived from the validation study of Rashidian et al. (2017). The parameters of the normal distribution are specified by the parameters of the approximate normal distribution of the bias parameter estimate, and the normal distribution is truncated at the limits of the 95% confidence interval of the bias parameter estimate, as shown in Table 4.9.

Table 4.9. Distribution of the bias parameters for sensitivity and specificity, using the truncated normal distribution

Bias parameter	Expectation (%)	Distribution parameters of the truncated normal distribution		
		Standard deviation	Minimum (%)	Maximum (%)
se_1	79	0.054 00	66	89
sp_1	83	0.033 77	76	90
se_0	68	0.076 96	50	82
sp_0	93	0.021 42	87	96

se_0, sensitivity for control participants; se_1, sensitivity for case participants; sp_0, specificity for control participants; sp_1, specificity for case participants.

Side Box 4.4. Multiple imputation (continued)

Example 4.25. Opium use and HNSCC – bias analysis for categorical data (continued)

With these choices, the error-adjusted odds ratio is 7.66. Because truncated normal distributions were chosen as the prior distributions, the result differs from that of the probabilistic bias analysis (where the error-adjusted odds ratio is 8.66), even though uninformative priors were chosen for p_1 and p_0.

To apply the MC-SIMEX method, one must calculate the unadjusted regression model, in this situation, a logistic regression model, and specify the misclassification matrices for case participants,

$$\boldsymbol{\Pi}_1 = \begin{pmatrix} 0.83 & 0.21 \\ 0.17 & 0.79 \end{pmatrix} \quad \text{(E4.9)}$$

and control participants,

$$\boldsymbol{\Pi}_0 = \begin{pmatrix} 0.93 & 0.32 \\ 0.07 & 0.68 \end{pmatrix} \quad \text{(E4.10)}$$

With the unadjusted regression model and the misclassification matrix, an error-adjusted odds ratio of 6.8 is obtained, using the R package simex (Lederer et al., 2019). (text continues on page 110)

Example 4.26. Non-differential outcome misclassification in studies of low-dose ionizing radiation

Linet et al. (2020) reviewed the potential for misclassification of leukaemia and all-cancer diagnosis in 26 studies of low-dose radiation exposure. False-negatives (underdiagnoses) were likely in only 2 of the 17 cancer incidence studies and 2 of the 9 mortality studies. False-positives (overdiagnoses) were likely in only one of the cancer incidence studies. Issues with the accuracy of the diagnoses were found in only two studies. (text continues on page 110)

Example 4.27. Non-differential outcome misclassification from underdiagnosis of prostate cancer

Bell et al. (2015) found the prevalence of incidental prostate cancer at autopsy to range from 5% (95% CI, 3–8%) at age < 30 years to 59% (95% CI, 48–71%) at age > 79 years. This may mean that undiagnosed prostate cancers are often classified as non-cases; this possibility is often overlooked in both cohort and case–control studies. (text continues on page 112)

Example 4.28. Non-differential outcome misclassification of tumour subtypes

Night shift work was seen to be more strongly associated with high-grade prostate cancer than with low-grade tumours (Papantoniou et al., 2015); however, there is evidence that, among proven cases of prostate cancer, detection of high-grade cancer has a sensitivity of 72% and a specificity of 92% upon initial diagnosis. If these errors are non-differential with respect to the exposure, then the expectation is that the association will be biased towards the null. (text continues on page 112)

Example 4.29. Differential outcome misclassification among firefighters

An increased risk of prostate cancer could be observed in studies of firefighters, because they are likely to undergo more medical screening than the general population used as the referent (DeBono et al., 2023). This was an important consideration in the *IARC Monographs* Working Group's determination that there was *limited* evidence for a causal association between occupational exposure as a firefighter and prostate cancer (IARC, 2023). (text continues on page 112)

Example 4.30. Sensitivity analysis for outcome misclassification

In a study of night shift work and prostate cancer, analyses were conducted excluding control participants who had not recently been screened for this cancer and who therefore had a greater likelihood of having undetected prostate cancer. The findings from this study were not altered, suggesting that the lack of an association between night shift work and prostate cancer in this study was not due to the inclusion of unrecognized cases of prostate cancer in the control group (Barul et al., 2019). (text continues below)

4.5 Summary

Errors in the measurement of both exposures and outcomes are potential sources of information bias in epidemiological studies. The errors for exposure measurement may be due to either misclassification (for a categorical classification) or mismeasurement (for a continuous measure). Unless exposure is measured prospectively, epidemiological studies of exposures associated with cancer risk are particularly prone to this source of bias, because many cancers have a long latency (time since first exposure) period (e.g. > 20 years), and therefore the relevant exposures may have occurred many years earlier. Misclassification or mismeasurement of cancer outcomes is less common but may occur when mortality data rather than incidence data are used, when case ascertainment is low (e.g. because of poor access to diagnostic health care), when a diagnostic test is used that has poor sensitivity and specificity (e.g. for prostate cancer), or because of changes in diagnostic categories over time (e.g. for mesothelioma or lymphatic and haematopoietic neoplasms).

Table 4.10 summarizes the expected direction of the bias for different types of error. If the errors in exposure measurement are random and non-differential with respect to disease status, the resulting information bias would be expected to be towards the null in studies with a binary (yes or no) exposure. However, the bias can be in either direction if the analysis includes more than two categories of exposure (e.g. high, medium, or low); in this situation,

Table 4.10. Summary of expected direction of bias in the effect estimate due to exposure misclassification and measurement error, and methods that may be used for correction or for assessing the potential magnitude of the biases using sensitivity analyses

Exposure metric	Error type	Expected direction of bias[a]	Methods for adjustment	Data needed for adjustment	Comments
Binary (yes or no)	Non-differential	Towards the null	Simple analysis	Simple 2 × 2 table of results; se and sp from a validation study	Assumptions can be made about se and sp if a validation study is not available.
	Differential	Either direction	Multidimensional analysis	Simple 2 × 2 table of results; range of plausible se and sp	The range of se and sp can be a plausible range chosen by the investigator.
			Probabilistic analysis	Simple 2 × 2 table of results; se and sp from a validation study; distribution of se and sp	Assumptions can be made about the bias parameters if data on se and sp are not available.
Multilevel	Non-differential or differential	Either direction	MC-SIMEX	Raw data; misclassification matrices from a validation study	
Continuous	Non-differential				
	Classical	Towards the null	Regression calibration	Data from a validation study	
	Linear	Either direction	Regression calibration	Data from a validation study	
	Berkson	Unbiased for linear models	No adjustment required		Non-linear models are generally close to unbiased if the outcome is rare. Berkson error is unbiased only if it is independent of other covariates.
	Differential	Either direction	Multiple imputation	Data from an internal validation study for case and non-case participants	

MC-SIMEX, simulation extrapolation for misclassification; se, sensitivity; sp, specificity.

[a] The expected direction of the bias is what is generally expected to be observed over a large number of trials or studies. An individual study finding may or may not be biased in the direction expected, because of random variation.

misclassification of exposure is most likely to result in overestimation of risk in an intermediate exposure category but underestimation in the highest exposure category, and there can even be a change in the direction of the slope across exposure categories under certain conditions (Dosemeci et al., 1990; Weinberg et al., 1994). Thus, categorization of a non-differentially misclassified continuous exposure variable can result in differential misclassification (Flegal et al., 1986). The bias can also be in either direction if the errors are differential with respect to disease.

For continuous measures, the effect of measurement error depends on the error structure, which could involve combinations of systematic error and random error following classical, linear, or Berkson error structures. These error structures could be additive, multiplicative, or mixed. Classical errors occur when there is an erroneous measurement method that gives the correct value on average but yields a somewhat different value each time it is applied, sometimes larger than and sometimes smaller than the true exposure. The bias arising from using an exposure measure that has classical

errors is expected to attenuate the slope of the exposure–response relation. A linear model describes an erroneous measurement method that, on average, does not give the correct value of the exposure (i.e. is biased). The effect of using an exposure measure with errors that are linear could be in either direction, depending on whether the expected value of the exposure is less than or greater than the true exposure. Finally, the Berkson error model is similar to a classical error model in having a mean of zero but, unlike in the classical error model, the error is not independent of the true value. Berkson errors are common in occupational studies where a group mean is used to describe the exposures of workers engaged in a particular job. Using exposure measurements that have a Berkson error structure does not generally bias the effect measures but does increase standard errors. It is noteworthy that a particular study may be subject to a combination of these three error types; in this situation, the direction of the bias may be difficult to predict.

Differential misclassification of exposure is a common concern in studies that rely on questionnaire data to assess exposure. This is a problem particularly in case–control studies, in which interviews are conducted after the case status is known. It is less often a concern in cohort studies, in which exposure information is generally assessed before the disease occurrence. Recall bias and interviewer bias can introduce differential misclassification of exposure. Blinding of the interviewers to the case status makes interview bias unlikely but will usually have little effect on recall bias. Interviews of proxies (e.g. next of kin) are often used in case–control studies where the case participants are deceased; this may result in differential information bias (e.g. if the proxies of deceased case participants have poorer knowledge of the case participants' exposures than the living control participants have of theirs). The effect of differential misclassification may be in either direction. Recall and interviewer biases are usually away from the null because case participants are more likely than healthy control participants to recall their exposures, and interviewers may be more likely to question case participants more deeply than control participants for their exposure histories. Proxy interviewees would generally be expected to be less likely than control participants to recall exposure, resulting in a bias towards the null.

There have been substantial developments in methods for assessing the magnitude of errors and adjusting for these biases. These methods, which are summarized in Table 4.10, may also be adapted for assessing and adjusting for errors in outcome classification. Some of these methods require the use of data from validation studies, in which the measurement method used in the study is compared with a gold standard. Frequently, results from validation studies may not be available to an *IARC Monographs* Working Group or other expert reviewers. However, a description of these methods is included, in the anticipation that more investigators will perform validation studies in the future. Sensitivity analyses can be conducted in most instances to estimate the magnitude of the error where assumptions are made about the sensitivity and specificity of the measurement methods. These methods can apply to situations where the errors are non-differential or differential with respect to exposure and can also be extended to include a range of plausible values of sensitivity and specificity. These simple methods can provide reviewers with some perspective on how large or small a true association might be. Biases for continuous measures of exposure can be corrected using regression calibration, using data from a validation study. Methods that require access to the raw study data (e.g. multiple imputation), which will not generally be available to an expert review group, are also discussed in Chapter 7.

References

Armstrong BG (1998). Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Environ Med. 55(10):651–6. doi:10.1136/oem.55.10.651 PMID:9930084

Arnold BF, Ercumen A, Benjamin-Chung J, Colford JMJ Jr (2016). Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies. Epidemiology. 27(5):637–41. doi:10.1097/EDE.0000000000000504 PMID:27182642

Barul C, Richard H, Parent ME (2019). Night-shift work and risk of prostate cancer: results from a Canadian case-control study, the Prostate Cancer and Environment Study. Am J Epidemiol. 188(10):1801–11. doi:10.1093/aje/kwz167 PMID:31360990

Baumgarten M, Siemiatycki J, Gibbs GW (1983). Validity of work histories obtained by interview for epidemiologic purposes. Am J Epidemiol. 118(4):583–91. doi:10.1093/oxfordjournals.aje.a113663 PMID:6637985

Bell KJ, Del Mar C, Wright G, Dickinson J, Glasziou P (2015). Prevalence of incidental prostate cancer: a systematic review of autopsy studies. Int J Cancer. 137(7):1749–57. doi:10.1002/ijc.29538 PMID:25821151

Berkson J (1950). Are there two regressions? J Am Stat Assoc. 45(250):164–80. doi:10.1080/01621459.1950.10483349

Blair A, Cantor KP, Zahm SH (1998). Non-Hodgkin's lymphoma and agricultural use of the insecticide lindane. Am J Ind Med. 33(1):82–7. doi:10.1002/(SICI)1097-0274(199801)33:1<82::AID-AJIM9>3.0.CO;2-Y PMID:9408531

Brown LM, Dosemeci M, Blair A, Burmeister L (1991). Comparability of data obtained from farmers and surrogate respondents on use of agricultural pesticides. Am J Epidemiol. 134(4):348–55. doi:10.1093/oxfordjournals.aje.a116096 PMID:1877595

Cardis E, Richardson L, Deltour I, Armstrong B, Feychting M, Johansen C, et al. (2007). The INTERPHONE study: design, epidemiological methods, and description of the study population. Eur J Epidemiol. 22(9):647–64. doi:10.1007/s10654-007-9152-z PMID:17636416

Chu H, Wang Z, Cole SR, Greenland S (2006). Sensitivity analysis of misclassification: a graphical and a Bayesian approach. Ann Epidemiol. 16(11):834–41. doi:10.1016/j.annepidem.2006.04.001 PMID:16843678

Cole SR, Chu H, Greenland S (2006). Multiple-imputation for measurement-error correction. Int J Epidemiol. 35(4):1074–81. doi:10.1093/ije/dyl097 PMID:16709616

Cook JR, Stefanski LA (1994). Simulation-extrapolation estimation in parametric measurement error models. J Am Stat Assoc. 89(428):1314–28. doi:10.1080/01621459.1994.10476871

Corbin M, Haslett S, Pearce N, Maule M, Greenland S (2017). A comparison of sensitivity-specificity imputation, direct imputation and fully Bayesian analysis to adjust for exposure misclassification when validation data are unavailable. Int J Epidemiol. 46(3):1063–72. doi:10.1093/ije/dyx027 PMID:28338966

Costas L, Infante-Rivard C, Zock JP, Van Tongeren M, Boffetta P, Cusson A, et al. (2015). Occupational exposure to endocrine disruptors and lymphoma risk in a multi-centric European study. Br J Cancer. 112(7):1251–6. doi:10.1038/bjc.2015.83 PMID:25742473

Daniels RD, Schubauer-Berigan MK (2005). Bias and uncertainty of penetrating photon dose measured by film dosemeters in an epidemiological study of US nuclear workers. Radiat Prot Dosimetry. 113(3):275–89. doi:10.1093/rpd/nch470 PMID:15769802

DeBono NL, Daniels RD, Beane Freeman LE, Graber JM, Hansen J, Teras LR, et al. (2023). Firefighting and cancer: a meta-analysis of cohort studies in the context of cancer hazard identification. Saf Health Work. 14(2):141–52. doi:10.1016/j.shaw.2023.02.003 PMID:37389311

Dosemeci M, Wacholder S, Lubin JH (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol. 132(4):746–8. doi:10.1093/oxfordjournals.aje.a115716 PMID:2403115

Edwards S, Slattery ML, Mori M, Berry TD, Caan BJ, Palmer P, et al. (1994). Objective system for interviewer performance evaluation for use in epidemiologic studies. Am J Epidemiol. 140(11):1020–8.doi:10.1093/oxfordjournals.aje.a117192 PMID:7985650

Eurostat (2022). Data browser. Employed persons working at nights as a percentage of the total employment, by sex, age and professional status (%). Available from: https://ec.europa.eu/eurostat/databrowser/view/LFSA_EWPNIG/default/table?lang=en.

Finkelstein MM (2000). Leukemia after exposure to benzene: temporal trends and implications for standards. Am J Ind Med. 38(1):1–7. doi:10.1002/1097-0274(200007)38:1<1::AID-AJIM1>3.0.CO;2-9 PMID:10861761

Flegal KM, Brownie C, Haas JD (1986). The effects of exposure misclassification on estimates of relative risk. Am J Epidemiol. 123(4):736–51. doi:10.1093/oxfordjournals.aje.a114294 PMID:3953551

Fox MP, Lash TL, Greenland S (2005). A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol. 34(6):1370–6. doi:10.1093/ije/dyi184 PMID:16172102

Fox MP, MacLehose RF, Lash TL (2021). Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer. doi:10.1007/978-3-030-82673-4

Fox MP, MacLehose RF, Lash TL (2023). SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. Int J Epidemiol. 52(5):1624–33. doi:10.1093/ije/dyad053 PMID:37141446

Freedman LS, Schatzkin A, Midthune D, Kipnis V (2011). Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst. 103(14):1086–92. doi:10.1093/jnci/djr189 PMID:21653922

Fritschi L, Erren TC, Glass DC, Girschik J, Thomson AK, Saunders C, et al. (2013). The association between different night shiftwork factors and breast cancer: a case-control study. Br J Cancer. 109(9):2472–80. doi:10.1038/bjc.2013.544 PMID:24022188

Fritschi L, Valérie Groß J, Wild U, Heyworth JS, Glass DC, Erren TC (2018). Shift work that involves circadian disruption and breast cancer: a first application of chronobiological theory and the consequent challenges. Occup Environ Med. 75(3):231–4. doi:10.1136/oemed-2017-104441 PMID:28775132

Ge C, Peters S, Olsson A, Portengen L, Schüz J, Almansa J, et al. (2020). Respirable crystalline silica exposure, smoking, and lung cancer subtype risks. A pooled analysis of case-control studies. Am J Respir Crit Care Med. 202(3):412–21. doi:10.1164/rccm.201910-1926OC PMID:32330394

Gilbert ES (2009). The impact of dosimetry uncertainties on dose-response analyses. Health Phys. 97(5):487–92. doi:10.1097/HP.0b013e3181adc3b1 PMID:19820458

Gilbert R, Martin RM, Donovan J, Lane JA, Hamdy F, Neal DE, et al. (2016). Misclassification of outcome in case-control studies: methods for sensitivity analysis. Stat Methods Med Res. 25(5):2377–93. doi:10.1177/0962280214523192 PMID:25217446

Glasziou PP, Irwig L, Heritier S, Simes RJ, Tonkin A; LIPID Study Investigators (2008). Monitoring cholesterol levels: measurement error or true change? Ann Intern Med. 148(9):656–61. doi:10.7326/0003-4819-148-9-200805060-00005 PMID:18458278

Goldsmith ES, Krebs EE, Ramirez MR, MacLehose RF (2023). Opioid-related mortality in United States death certificate data: a quantitative bias analysis with expert elicitation of bias parameters. Epidemiology. 34(3):421–9. doi:10.1097/EDE.0000000000001600 PMID:36735892

Greenland S (1988). Variance estimation for epidemiologic effect estimates under misclassification. Stat Med. 7(7):745–57. doi:10.1002/sim.4780070704 PMID:3043623

Greenland S (2009). Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. Int J Epidemiol. 38(6):1662–73. doi:10.1093/ije/dyp278 PMID:19744933

Greenland S, Robins JM (1985). Confounding and misclassification. Am J Epidemiol. 122(3):495–506. doi:10.1093/oxfordjournals.aje.a114131 PMID:4025298

Hansen J (2001). Increased breast cancer risk among women who work predominantly at night. Epidemiology. 12(1):74–7. doi:10.1097/00001648-200101000-00013 PMID:11138824

Heid IM, Lamina C, Küchenhoff H, Fischer G, Klopp N, Kolz M, et al. (2008). Estimating the single nucleotide polymorphism genotype misclassification from routine double measurements in a large epidemiologic sample. Am J Epidemiol. 168(8):878–89. doi:10.1093/aje/kwn208 PMID:18791193

Hernán MA, Cole SR (2009). Invited commentary: causal diagrams and measurement bias. Am J Epidemiol. 170(8):959–62, discussion 963–4. doi:10.1093/aje/kwp293 PMID:19755635

Hoffmann S, Laurier D, Rage E, Guihenneuc C, Ancelet S (2018). Shared and unshared exposure measurement error in occupational cohort studies and their effects on statistical inference in proportional hazards models. PLoS One. 13(2):e0190792. doi:10.1371/journal.pone.0190792 PMID:29408862

IARC (2018). Benzene. IARC Monogr Eval Carcinog Risks Hum. 120:1–301. Available from: https://publications.iarc.who.int/576 PMID:31769947

IARC (2020). Night shift work. IARC Monogr Identif Carcinog Hazard Hum. 124:1–371. Available from: https://publications.iarc.who.int/593 PMID:33656825

IARC (2021). Opium consumption. IARC Monogr Identif Carcinog Hazard Hum. 126:1–253. Available from: https://publications.iarc.who.int/600 PMID:36395294

IARC (2023). Occupational exposure as a firefighter. IARC Monogr Identif Carcinog Hazard Hum. 132:1–728. Available from: https://publications.iarc.who.int/615 PMID:37963216

INTERPHONE Study Group (2010). Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. Int J Epidemiol. 39(3):675–94. doi:10.1093/ije/dyq079 PMID:20483835

Johnson CY, Flanders WD, Strickland MJ, Honein MA, Howards PP (2014). Potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential binary exposure misclassification. Epidemiology. 25(6):902–9. doi:10.1097/EDE.0000000000000166 PMID:25120106

Jurek AM, Greenland S, Maldonado G, Church TR (2005). Proper interpretation of non-differential misclassification effects: expectations vs observations. Int J Epidemiol. 34(3):680–7. doi:10.1093/ije/dyi060 PMID:15802377

Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, et al. (2020). STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1-Basic theory and simple methods of adjustment. Stat Med. 39(16):2197–231. doi:10.1002/sim.8532 PMID:32246539

Kipnis V, Izmirlian G (2002). The impact of categorization of continuous exposure measured with error [abstract]. Am J Epidemiol. 155(11):S28.

Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, et al. (2003). Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol. 158(1):14–21, discussion 22–6. doi:10.1093/aje/kwg091 PMID:12835281

Koenig W, Sund M, Fröhlich M, Löwel H, Hutchinson WL, Pepys MB (2003). Refinement of the association of serum C-reactive protein concentration and coronary heart disease risk by correction for within-subject variation over time: the MONICA Augsburg studies, 1984 and 1987. Am J Epidemiol. 158(4):357–64. doi:10.1093/aje/kwg135 PMID:12915501

Kromhout H, Vermeulen R (2001). Application of job exposure matrices in studies of the general population; some clues to their importance. Eur Respir Rev. 11(80):80–90.

Küchenhoff H, Mwalili SM, Lesaffre E (2006). A general method for dealing with misclassification in regression: the misclassification SIMEX. Biometrics. 62(1):85–96. doi:10.1111/j.1541-0420.2005.00396.x PMID:16542233

Larsson SC, Rafter J, Holmberg L, Bergkvist L, Wolk A (2005). Red meat consumption and risk of cancers of the proximal colon, distal colon and rectum: the Swedish Mammography Cohort. Int J Cancer. 113(5):829–34. doi:10.1002/ijc.20658 PMID:15499619

Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S (2014). Good practices for quantitative bias analysis. Int J Epidemiol. 43(6):1969–85. doi:10.1093/ije/dyu149 PMID:25080530

Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ (2021). Modern epidemiology. 4th ed. Philadelphia (PA), USA: Wolters Kluwer.

Lawlor DA, Tilling K, Davey Smith G (2016). Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866–86. PMID:28108528

Lederer W, Seibold H, Küchenhoff H, Lawrence C, Brøndum RF (2019). Simex: SIMEX- and MCSIMEX-algorithm for measurement error models. R package, version 1.8. Available from: https://CRAN.R-project.org/package=simex.

Linet MS, Schubauer-Berigan MK, Berrington de González A (2020). Outcome assessment in epidemiologic studies of low-dose radiation exposure and cancer risks: sources, level of ascertainment, and misclassification. J Natl Cancer Inst Monogr. 2020(56):154–75. doi:10.1093/jncimonographs/lgaa007 PMID:32657350

Lipsitch M, Tchetgen Tchetgen E, Cohen T (2010). Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 21(3):383–8. doi:10.1097/EDE.0b013e3181d61eeb PMID:20335814

Loken E, Gelman A (2017). Measurement error and the replication crisis. Science. 355(6325):584–5. doi:10.1126/science.aal3618 PMID:28183939

MacLehose RF, Gustafson P (2012). Is probabilistic bias analysis approximately Bayesian? Epidemiology. 23(1):151–8. doi:10.1097/EDE.0b013e31823b539c PMID:22157311

Mohebbi E, Hadji M, Rashidian H, Rezaianzadeh A, Marzban M, Haghdoost AA, et al. (2021). Opium use and the risk of head and neck squamous cell carcinoma. Int J Cancer. 148(5):1066–76. doi:10.1016/j.juro.2016.08.095 PMID:27582435

Momoli F, Siemiatycki J, McBride ML, Parent MÉ, Richardson L, Bedard D, et al. (2017). Probabilistic multiple-bias modeling applied to the Canadian data from the Interphone Study of mobile phone use and risk of glioma, meningioma, acoustic neuroma, and parotid gland tumors. Am J Epidemiol. 186(7):885–93. doi:10.1093/aje/kwx157 PMID:28535174

Papantoniou K, Castaño-Vinyals G, Espinosa A, Aragonés N, Pérez-Gómez B, Ardanaz E, et al. (2016). Breast cancer risk and night shift work in a case-control study in a Spanish population. Eur J Epidemiol. 31(9):867–78. doi:10.1007/s10654-015-0073-y PMID:26205167

Papantoniou K, Castaño-Vinyals G, Espinosa A, Aragonés N, Pérez-Gómez B, Burgos J, et al. (2015). Night shift work, chronotype and prostate cancer risk in the MCC-Spain case-control study. Int J Cancer. 137(5):1147–57. doi:10.1002/ijc.29400 PMID:25530021

Rashidian H, Hadji M, Marzban M, Gholipour M, Rahimi-Movaghar A, Kamangar F, et al. (2017). Sensitivity of self-reported opioid use in case-control studies: healthy individuals versus hospitalized patients. PLoS One. 12(8):e0183017. doi:10.1371/journal.pone.0183017 PMID:28854228

Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 5(6B):1113–24. doi:10.1079/PHN2002394 PMID:12639222

Rosner B, Spiegelman D, Willett WC (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol. 132(4):734–45. doi:10.1093/oxfordjournals.aje.a115715 PMID:2403114

Rosner B, Willett WC, Spiegelman D (1989). Correction of logistic regression relative risk estimates for systematic within-person measurement error. Stat Med. 8:1051–69. doi:10.1002/sim.4780080905 PMID:2799131

Schubauer-Berigan MK, Daniels RD, Fleming DA, Markey AM, Couch JR, Ahrenholz SH, et al. (2007a). Risk of chronic myeloid and acute leukemia mortality after exposure to ionizing radiation among workers at four US nuclear weapons facilities and a nuclear naval shipyard. Radiat Res. 167(2):222–32. doi:10.1667/RR0724.1 PMID:17390730

Schubauer-Berigan MK, Daniels RD, Fleming DA, Markey AM, Couch JR, Ahrenholz SH, et al. (2007b). Chronic lymphocytic leukaemia and radiation: findings among workers at five US nuclear facilities and a review of the recent literature. Br J Haematol. 139(5):799–808. doi:10.1111/j.1365-2141.2007.06843.x PMID:17922878

Shakeri R, Kamangar F, Nasrollahzadeh D, Nouraie M, Khademi H, Etemadi A, et al. (2012). Is opium a real risk factor for esophageal cancer or just a methodological artifact? Hospital and neighborhood controls in case-control studies. PLoS One. 7(3):e32711. doi:10.1371/journal.pone.0032711 PMID:22396792

Shaw PA, Deffner V, Keogh RH, Tooze JA, Dodd KW, Küchenhoff H, et al.; Measurement Error and Misclassification Topic Group (TG4) of the STRATOS Initiative (2018). Epidemiologic analyses with error-prone exposures: review of current practice and recommendations. Ann Epidemiol. 28(11):821–8. doi:10.1016/j.annepidem.2018.09.001 PMID:30316629

Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Keogh RH, et al. (2020). STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 2-More complex methods of adjustment and advanced topics. Stat Med. 39(16):2232–63. doi:10.1002/sim.8531 PMID:32246531

Slate EH, Bandyopadhyay D (2009). An investigation of the MC-SIMEX method with application to measurement error in periodontal outcomes. Stat Med. 28(28):3523–38. doi:10.1002/sim.3656 PMID:19902495

Smith AH, Pearce NE, Callas PW (1988). Cancer case-control studies with other cancers as controls. Int J Epidemiol. 17(2):298–306. doi:10.1093/ije/17.2.298 PMID:3042650

Spearman C (1904). The proof and measurement of association between two things. Am J Psychol. 15(1):72–101. doi:10.2307/1412159 PMID:3322052

Stayner L, Vrijheid M, Cardis E, Stram DO, Deltour I, Gilbert SJ, et al. (2007). A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers. Radiat Res. 168(6):757–63. doi:10.1667/RR0677.1 PMID:18088178

Stewart PA, Lees PS, Francis M (1996). Quantification of historical exposures in occupational cohort studies. Scand J Work Environ Health. 22(6):405–14. doi:10.5271/sjweh.161 PMID:9000307

Stram DO, Kopecky KJ (2003). Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: some observations. Radiat Res. 160(4):408–17. doi:10.1667/3046 PMID:12968933

Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. (2003). Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. Am J Epidemiol. 158(1):1–13. doi:10.1093/aje/kwg092 PMID:12835280

Thierry-Chef I, Marshall M, Fix JJ, Bermann F, Gilbert ES, Hacker C, et al. (2007). The 15-Country Collaborative Study of Cancer Risk among Radiation Workers in the Nuclear Industry: study of errors in dosimetry. Radiat Res.167(4):380–95.doi:10.1667/RR0552.1 PMID:17388692

Thierry-Chef I, Richardson DB, Daniels RD, Gillies M, Hamra GB, Haylock R, et al.; INWORKS Consortium (2015). Dose estimation for a study of nuclear workers in France, the United Kingdom and the United States of America: methods for the International Nuclear Workers Study (INWORKS). Radiat Res. 183(6):632–42. doi:10.1667/RR14006.1 PMID:26010707

Villanueva CM, Silverman DT, Malats N, Tardon A, Garcia-Closas R, Serra C, et al. (2009). Determinants of quality of interview and impact on risk estimates in a case-control study of bladder cancer. Am J Epidemiol. 170(2):237–43. doi:10.1093/aje/kwp136 PMID:19478234

Vrijheid M, Armstrong BK, Bédard D, Brown J, Deltour I, Iavarone I, et al. (2009). Recall bias in the assessment of exposure to mobile phones. J Expo Sci Environ Epidemiol. 19(4):369–81. doi:10.1038/jes.2008.27 PMID:18493271

Vrijheid M, Cardis E, Armstrong BK, Auvinen A, Berg G, Blaasaas KG, et al.; Interphone Study Group (2006). Validation of short-term recall of mobile phone use for the Interphone study. Occup Environ Med. 63(4):237–43. doi:10.1136/oem.2004.019281 PMID:16556742

Wacholder S (1995). When measurement errors correlate with truth: surprising effects of nondifferential misclassification. Epidemiology. 6(2):157–61.doi:10.1097/00001648-199503000-00012 PMID:7742402

Weinberg CR, Umbach DM, Greenland S (1994). When will nondifferential misclassification of an exposure preserve the direction of a trend? Am J Epidemiol. 140(6):565–71. doi:10.1093/oxfordjournals.aje.a117283 PMID:8067350

Yland JJ, Wesselink AK, Lash TL, Fox MP (2022). Misconceptions about the direction of bias from nondifferential misclassification. Am J Epidemiol. 191(8):1485–95. doi:10.1093/aje/kwac035 PMID:35231925