Chapter 7. Study reporting considerations to facilitate quantitative bias assessment with access to original data

CHAPTER 7.

Study reporting considerations to facilitate quantitative bias assessment with access to original data

Lin Fritschi, Terry Boyle, Brigid M. Lynch, Scott Weichenthal, and Irina Guseva Canu

7.1 Introduction

The previous chapters in this book introduced and explained common biases in epidemiological studies, as well as methods for quantitative bias analysis that are suitable for use in systematic reviews and the hazard identification process.

In contrast, this chapter is aimed at researchers who have access to individual-level data and wish to undertake quantitative bias analysis themselves or to facilitate the inclusion of their study results in systematic reviews and hazard identifications. The goal here is to provide researchers with clear information on what they need to report to facilitate the bias assessment process, whether the bias assessment is carried out by the study team themselves or their study is being examined by systematic reviewers and hazard assessors.

As with the rest of this book, this chapter focuses on confounding, information bias (measurement error and misclassification of exposures and outcomes), and selection bias. For each type of bias, a brief description is first provided of how the bias may arise in epidemiological studies; this is followed, in some cases, by examples to illustrate how quantitative bias analyses can be conducted when individual-level information is available to study authors. To avoid duplication, readers are referred to the relevant sections of Chapters 2–5 for more details about the methods and biases discussed in this chapter.

A special point has been made of tabulating the specific information that must be reported to facilitate each type of bias assessment. The required parameters are described, as well as their use in the bias assessment process. Some statistical packages that can be used to perform the quantitative bias analysis are also mentioned.

Importantly, this chapter does not discuss ways in which bias can be addressed by improvements to study design. Specifically, those situations are presented in which researchers do not have the option to alter the design of the study or to collect further data. There are several common scenarios where this may occur. The first scenario is when a researcher is analysing data from an existing study, such as a large cohort study or case–control study. This is most likely to be the situation when a new hypothesis is investigated using

data from an existing study that has been under way for many years or for which data collection has been completed, or when the follow-up of an existing cohort is extended. The second scenario is in a researcher's own study, where the depth and accuracy of data on important variables cannot be improved (e.g. if using existing medical records for assessing exposures, outcomes, or confounders) or where the study has been completed and the study design cannot be changed. A third scenario is in the analysis of data from large consortia in which individual studies are pooled or combined, and where the data from the individual studies may have different biases.

Moreover, this chapter does not take the approach of the many checklists and tools that have been developed to assess whether there is a risk of bias. In a review, Wang et al. (2019) identified 62 tools aimed at assessing the risk of bias in observational studies of exposures. Almost half of the tools that were reviewed enabled the calculation of a quality score, although Wang et al. (2019) questioned whether these scores were useful. Although these types of tool may be useful for authors or reviewers to provide an initial examination of a study to determine whether there is a risk of bias, none of them is able to provide a quantitative estimate of the direction or magnitude of the bias (Savitz et al., 2019).

Finally, another goal of this chapter is to encourage researchers to replace qualitative comments on the role of bias in their studies with quantitative estimates based on formal bias analysis. Too often, discussion sections of papers contain general statements in which authors describe the study's limitations qualitatively. The authors may estimate the assumed direction and sometimes provide a qualitative description of the effect of errors, such as selection bias, confounding, or information bias, based primarily on their knowledge of the field and of their own study. However, as discussed by Lash et al. (2021), human reasoning under uncertainty is well known to be fallible and to be biased by previous experience, by conflicts of interest, and also by the tendency to favour exposure effects over systematic errors as an explanation for observed associations. It is hoped that the information provided in this chapter will assist researchers to assess the direction and quantify the magnitude of systematic errors in their studies and to report the information required to facilitate the development of systematic reviews and hazard identifications.

When considering biases in observational epidemiology studies, it may be useful to conceptualize a target trial. While a detailed examination of target trials is beyond the scope of this book, some conceptual background is provided in Side Box 7.1.

Section 7.2 outlines the reporting considerations to facilitate graphical analysis of the biases in a study. Sections 7.3, 7.4, and 7.5 address considerations to facilitate quantitative bias analyses related to confounding, exposure misclassification or measurement error, and selection bias. Each section includes summary tables highlighting important reporting considerations and worked examples to illustrate how this information can be used to support bias assessment.

7.2 Reporting considerations to aid graphical approaches to identify biases

A detailed description of how directed acyclic graphs (DAGs) can be used in the hazard identification process is provided in Chapter 2, including definitions, components, interpretation, and their application in identifying potential biases in epidemiological analyses. This section focuses on reporting principles that can be implemented in constructing and presenting DAGs to facilitate bias assessment. These principles can be applied at the study design or analysis stages, or both (i.e. to explicitly describe assumptions being made with respect to the data-generation process) or in evaluating existing scientific evidence (i.e. by reconstructing the implied relations between exposures, outcomes, and covariates to evaluate potential sources of bias that were not addressed in the initial analysis).

Briefly, DAGs provide a formal mechanism for investigators to explicitly outline assumptions made regarding structural relations between exposures, outcomes, and covariates, both measured and unmeasured (e.g. confounders, intermediates, and collider variables), relevant to a given question. Through this process, DAGs also play a crucial role in enabling the identification of potential biases (e.g. confounding or selection bias; see Chapter 2) that must be addressed in estimating the causal relation between an exposure and an outcome. With respect to reporting, Tennant et al. (2021) list eight recommendations to improve the transparency and

Side Box 7.1. Target trials

Target trial approaches, which anchor causal assumptions to study design and analysis (Hernán, 2016; Hernán and Robins, 2020), can improve causal inference in observational studies and address common biases. Target trial emulation applies the principles of randomized controlled trials to observational data analysis. This is done by describing the protocol of an ideal randomized controlled trial that could be used to answer the research question of interest. The next step is to determine whether the research question can be identified and the outcomes estimated using observational data. Of course, there are always challenges when drawing causal inferences from observational studies, because of the pervasiveness of biases; exchangeability (i.e. an absence of confounding) cannot be guaranteed with non-randomized data. Furthermore, the target trial construct can be challenging to adapt to most occupational and environmental exposures that are typically the subject of *IARC Monographs* evaluations and in which exposure is protracted and latency is very long (Steenland et al., 2020). Nonetheless, the target trial approach can be a useful framework when carefully considering how to clearly articulate the causal effect to be estimated and biases that may affect the analysis (flagging the need for statistical methods to address these biases). Causal inference is improved by being transparent about causal assumptions, acknowledging uncertainties in the interpretation of causal effects, and striving to obtain the least-biased effect estimate within one's means (Hernán, 2016; Moreno-Betancur, 2021). Readers are referred to Hernán and Robins (2020, Chapter 22) for a detailed description of how to emulate a target trial.

In terms of reporting, authors are encouraged to describe their protocol components. This involves clearly defining the research question (the causal effect of interest), eligibility criteria, intervention (or exposure) characteristics and implementation, follow-up period, outcome of interest, and statistical analysis (specifying intention-to-treat or per-protocol effects). (text continues on page 177)

utility of DAGs in identifying potential biases; these recommendations can be summarized as follows.

(i) Clearly state the relations being focused on and the estimands of interest.
- Be clear about the exposures and outcomes of interest, including the level at which exposures are measured (e.g. environmental concentrations, personal exposures, biomarker concentrations).

(ii) A DAG should be presented for each focal relation and estimand of interest.
- Report a DAG for each causal relation under investigation.
- Online resources are available to support the construction of DAGs (e.g. DAGitty, Textor et al., 2016), and an R package is also available.

(iii) All relevant variables should be included in DAGs, even where direct measurements are unavailable.
- Include all possible confounding variables in the DAG, even those that were not measured. As described previously, DAGs can also be used to identify or describe possible sources of selection bias and measurement error, if these are a concern.
- It is useful to indicate in the DAG any variables that were not measured (e.g. using a different shape), to highlight potential sources of residual confounding.
- In some situations, many possible confounders may exist; including them all in the DAG can lead to cluttered and confusing diagrams. To avoid this, start by reporting only the most important confounders in the DAG (i.e. those that are expected to have an important impact on the hazard identification process). However, it is important to note that one's intuition about which are the most important variables can be wrong, and exclusion of variables should be justified.

(iv) Variables should be visually arranged so that all constituent arcs flow in the same direction.
- DAGs are easier to interpret when the constituent variables are arranged in a manner that clearly reflects the passage of time (i.e. exposure before outcome), with arcs flowing in the same direction (i.e. from left to right or from top to bottom).

(v) The omission of arrows and nodes should be carefully considered and justified with theory or evidence.
  - Omitting an arrow from one node to another implies no causal effect of one on the other.
  - This is a stronger assumption than including an arrow from one node to another (which can take any sign or magnitude, including a very small effect).

(vi) The DAG-implied adjustment sets for the estimands of interest should be clearly stated.
  - After the DAG is constructed, be clear about what it implies about the necessary adjustment set, including variables that may be missing because they were not measured.

(vii) Risk estimates obtained from the DAG-implied adjustment sets should be reported.
  - When the DAG-implied adjustment set has been identified, use it in the analysis and report the results. If some variables are missing (i.e. because they were not measured), it should be stated that the analysis is not based on the DAG-implied adjustment set.
  - Quantitative bias analysis can be used to estimate the potential impact of unmeasured confounders caused by missing variables (Lash et al., 2021), as described in Section 3.3.4.
  - For hazard identification, it can be helpful to report a minimally adjusted model to assess the extent of bias from the selected set of confounders, as described in Section 3.2.3. However, this should be interpreted with caution because other factors (e.g. measurement error for variables identified as confounders) will influence differences between adjusted and minimally adjusted models.

(viii) Alternative adjustment sets should be justified and reported separately.
  - If more than one adjustment set is used (including unadjusted models), these should be clearly justified, and the results should be reported separately from the DAG-implied adjustment set.

Hypothetical scenarios of DAG reporting are provided in Examples 7.1 and 7.2.

7.3 Confounding

A limitation of observational studies is that they are prone to the risk of residual or unmeasured confounding, which can lead to biased estimates of the effect of the exposure of interest (VanderWeele, 2019). A detailed description of confounding and how this affects causal estimates in epidemiological studies is provided in Chapter 3. Researchers try to minimize confounding by using methods related to study design (e.g. randomization, restriction, matching) or, after completion of data collection, by using multivariable analysis or stratification.

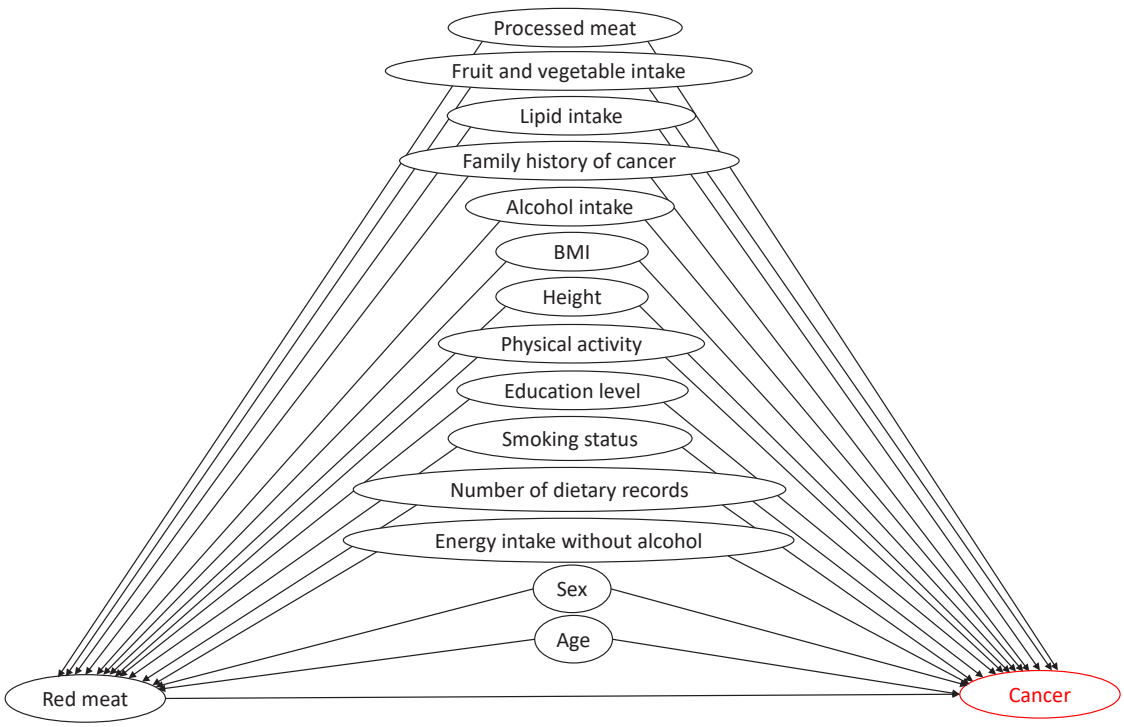
Chapter 3 discusses how to evaluate the adequacy of control for confounding in observational studies of cancer risk. This section focuses on controlling for confounding in secondary data analyses, i.e. when analysing data from case–control or cohort studies that have already been designed and conducted, including analyses using pooled data from large international consortia of these studies. For this purpose, it is assumed that the research questions addressed in secondary data analyses are causal ones (as opposed to descriptive or predictive questions).

In observational studies, it is only possible to attempt to emulate randomized experiments. For an observational study to emulate a randomized experiment, three assumptions must be satisfied (Shiba and Kawahara, 2021): conditional exchangeability (exposed and unexposed individuals are exchangeable within strata of the combinations of covariate values, i.e. there are no unmeasured confounders that are a common cause of both exposure and outcome); positivity (exposed and unexposed individuals are present within all combinations of covariate values); and consistency (the exposure is sufficiently well defined and has no variations that could alter the outcome). Identifying, measuring, and adjusting for confounders is crucial for the conditional exchangeability assumption (although the assumptions are interrelated). Note that in observational studies, one can never be sure what the true conditional randomization probability is (i.e. the likelihood of an outcome occurring, based on the occurrence of a previous outcome). The issue of residual and unmeasured confounding will always remain in observational studies (Hernán and Robins, 2020), but this section highlights methods to evaluate the direction and magnitude of uncontrolled confounding to help gauge how problematic it is likely to be.

Example 7.1. Red meat consumption and cancer

Diallo et al. (2018) examined the relation between red meat intake and cancer risk. Red meat intake was estimated through dietary records, and several different cancer outcomes were examined. Covariates identified as possible confounders in models for all cancers included age, sex, energy intake without alcohol, number of 24-hour dietary records, smoking status, education level, physical activity, height, body mass index, alcohol intake, family history of cancer, lipid intake, intake of fruits and vegetables, and intake of processed meat. This adjustment set implies the DAG in Fig. 7.1.

Fig. 7.1. Directed acyclic graph for red meat consumption and cancer. BMI, body mass index.

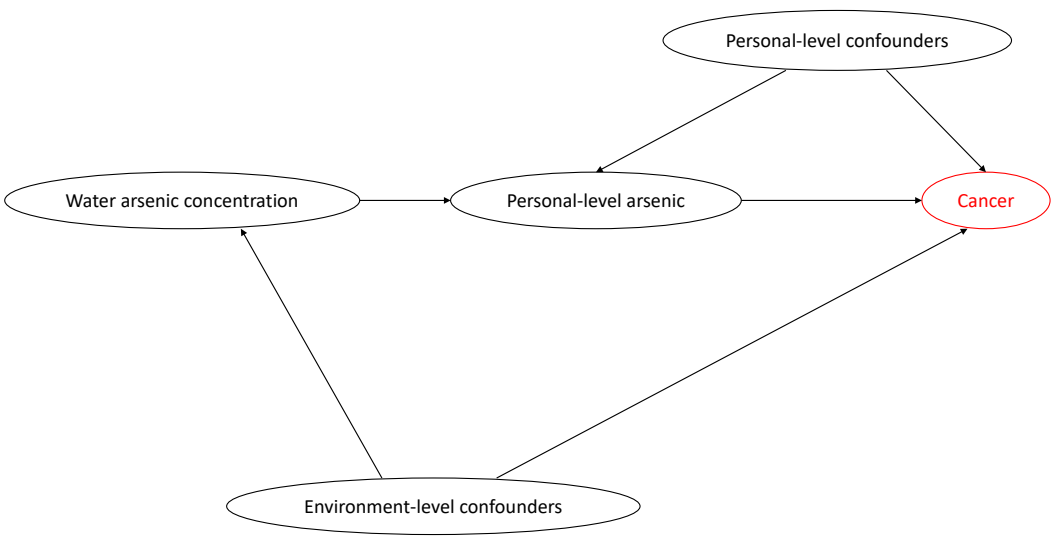


Clearly, many of the variables shown in Fig. 7.1 are likely to be important confounders (e.g. family history of cancer, smoking status). However, some of the variables included as confounders might be debatable (e.g. height), and an alternative adjustment set could be examined (e.g. by excluding height or other questionable confounding variables included in the analysis) to evaluate the impact of excluding those variables from the analysis. (text continues on page 179)

Example 7.2. Water arsenic concentration and cancer

Consider a hypothetical study where the agent of interest is water concentration of arsenic and not personal exposure to arsenic. This example is interesting because, for the purposes of reporting, it is important to differentiate between the levels at which exposure is measured (i.e. personal) and a more proxy level (e.g. environmental concentrations), because the set of potential confounders of the environmental concentration–outcome relation will probably differ from the set of potential confounders of the personal exposure–outcome relation (Weisskopf and Webster, 2017). Specifically, the association between the outcome and exposures measured at the personal level is more susceptible to confounding by individual-level factors (e.g. personal behaviours, such as diet or smoking), which can be difficult to measure and hard to control for in an analysis. For example, individual-level smoking is probably an important confounder of the relation between personal exposure to arsenic and cancer (because smoking is a cause of personal exposure to arsenic and smoking causes cancer) but is probably not an important confounder of the relation between water arsenic concentration and cancer incidence (because individual-level smoking is not a cause of arsenic in drinking-water). Alternatively, regional-level socioeconomic status (by postal code, county, etc.) may be an important confounder of the environmental concentration–outcome relation if areas with lower socioeconomic status have a higher incidence of cancer and have higher levels of arsenic in the water (e.g. because of a higher proportion of well-water use in rural areas with lower socioeconomic status). In addition, in retrospective studies, personal-level exposure measurements (e.g. biomarkers) could also be subject to reverse causation if the disease under investigation alters biomarker levels (see Chapter 5 for the issue of reverse causation). Fig. 7.2 is a generic DAG that highlights the distinction between confounders at the personal level and more-proxy-level confounders (e.g. environmental concentrations); it is important to think carefully about the variables that are likely to be present in each group and which variables need to be included in the analysis, based on the exposure of interest. A more thorough discussion of the trade-offs between personal and proxy-level exposures is given by Weisskopf and Webster (2017). (text continues on page 179)

Fig. 7.2. Distinguishing between personal-level confounders and more-proxy-level confounders (here, water arsenic concentration).



7.3.1 Reporting considerations to facilitate methods to assess confounding

Table 7.1 lists the important elements that should be reported to facilitate use of the tools described in Chapter 3 to assess bias from uncontrolled or residual confounding in the published literature. Researchers could provide this information in published studies to enable bias appraisal by themselves or reviewers of their work, as described here and in Chapter 3.

7.3.2 Methods of confounder selection

Confounders should be identified a priori, using a DAG, and documented in a statistical analysis plan. Contemporary epidemiological methods suggest that confounder selection should be based on sufficient knowledge of the relevant causal structures, and that the temporal relations of variables should be considered (VanderWeele, 2019). The use of DAGs for this purpose is discussed extensively in Chapter 2 and Section 7.2. The construction of DAGs can also help researchers consider which variable in a dataset best represents the confounder of interest.

Data-driven covariate selection – for example, forward or backward stepwise selection, examining *P* in bivariate analysis with either exposure or outcome, or examining a change in effect estimate after the addition or removal of a covariate – is not recommended (Greenland and Pearce, 2015). As a historical example, in a cohort study of the consumption of red and processed meats and colorectal cancer, English et al. (2004) stated, "Sex, country of birth, and energy intake (kJ/d) were included in all models. Other potential confounding variables were included in all the definitive analyses if they changed the hazard ratios of any of the meat consumption variables for either colon or rectal cancer by at least 5%." These methods do not consider the underlying causal structure, and it is not possible to determine whether covariates are confounders, mediators, colliders, or ancestors or descendants of other variables when using these data-driven approaches. When adjusting for covariates that are not true confounders, there is a risk of generating biased estimates.

As noted in Chapter 1, the *IARC Monographs* review process assigns greater weight to studies that adjust appropriately for confounding factors. Studies with insufficient adjustment are either given less weight or excluded from a review, depending on the number of studies available for a particular cancer site. Consideration of the method of confounder selection should also be part of this evaluation

Table 7.1. Essential information that is needed to inform assessment of bias from counfounding

Method to assess confounding	Data needed	More details
Negative control outcomes (NCOs)	Identification of NCO that is related to the confounder but not to the exposure[a] Reported results of the exposure–NCO association (as well as the main result of the exposure–disease[b] association)	Sections 3.3.2(a), 7.3.3(a)
Negative control exposures (NCEs)	Identification of NCE that is related to the confounder but is not a cause of disease Reported result for the NCE–disease association, adjusted for the exposure (or the NCE–disease association within a stratum of the exposure)	Sections 3.3.2(b), 7.3.3(b)
Bias analysis (e.g. indirect confounder adjustment)	For bounding, report a value for the probable magnitude of the association of the confounder with the disease in the population under study and of the confounder with the exposure For quantitative bias assessment, also report the prevalence of the confounder among unexposed (p_0) and exposed (p_1) individuals and information on the association between the confounder (e.g. smoking) and the exposure	Sections 3.3.4, 7.3.3(c)
Internal reference groups	Data on all exposure groups, including unexposed groups	Section 7.3.3(c)
External reference groups	Data on exposure and disease in external population used as reference	Section 7.3.3(c)
Duration of exposure	Dates of start and end of the exposure	Section 7.3.3(c)
g-methods	Data to enable simulation of the natural course of the disease with no intervention	Section 7.3.3(d)

[a] Exposure of primary interest.
[b] Disease of primary interest.

of evidence, given the potential for incorrect adjustment to introduce bias, as described in Section 3.2.3.

7.3.3 Addressing unmeasured and residual confounding

Commonly, bias in observational studies comes from confounders that are unmeasured or poorly measured (VanderWeele, 2019). Ideally, sensitivity analyses can be conducted to explore biases, including unmeasured and residual confounding; this can help with the interpretation of results and in avoiding the misapplication of study findings (Lash et al., 2014). There are several ways in which this can be done. Methods that do not require access to individual-level data include consideration of transportability of causal relations between studies (see Section 5.1), triangulation (see Section 3.3.3), and bounding and bias adjustment in sensitivity analyses (see Section 3.3.4).

When researchers have access to individual-level data, additional methods can be used to estimate the effects of unmeasured and residual confounding, including negative control outcomes (NCOs), negative control exposures (NCEs), and indirect control methods. These are described in more detail here.

(a) NCOs to address confounding

A detailed discussion of this method is presented in Section 3.3.2(a). The potential for confounding may be examined using an NCO (Lipsitch et al., 2010). This approach involves examining the association between the exposure of interest (the potential hazard) and another outcome that has the following characteristics: (i) it is caused by the hypothesized confounding factors, and (ii) it is not caused by the exposure (Example 7.3). If the association between the exposure of interest and the (implausible) NCO is of similar magnitude to the association between the exposure and the primary (plausible) outcome, this implies that the apparent association between the exposure and the primary outcome results from pervasive confounding. Researchers using NCOs must explicitly report how the selected NCOs meet these two conditions.

(b) NCEs to address confounding

An NCE approach is conceptually similar to the NCO method (see Section 3.3.2(b)), but here an alternative (implausible) exposure–outcome analysis is conducted. This method involves examining the association of a site-specific cancer outcome of interest with another exposure variable that has the following characteristics: (i) it is associated with the hypothesized confounding factors, and (ii) it is not a cause of the site-specific cancer outcome (Example 7.4). Researchers using an NCE must state how it meets these two conditions.

(c) Indirect methods to control for confounding

In some cohort and case–control studies, data on potentially important confounding factors may be missing. This applies particularly to retrospective studies of occupational exposures, where there was no unexposed group (Axelson and Steenland, 1988).

Example 7.3. Use of a negative control outcome in a study of hypertension and cancer

In a Mendelian randomization study, Chan et al. (2021) examined genetically predicted blood pressure and the risk of total and site-specific cancers. Single nucleotide polymorphisms (SNPs) that map to genes associated with systolic and diastolic blood pressure were identified in a genome-wide association study conducted using data obtained from the UK Biobank. These SNPs were used together to examine their collective relation with 17 site-specific cancers using data from a meta-analysis of the UK Biobank and the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging. Findings were validated using data from three international consortia (the Breast Cancer Association Consortium, the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome Consortium, and the International Lung Cancer Consortium). Asthma was used as an NCO, because blood pressure is unrelated to asthma but they share similar confounders (e.g. tobacco smoke exposure, obesity, physical inactivity). Systolic and diastolic blood pressure were not associated with risk of total or most site-specific cancers, and consistent findings were obtained from the validation samples. There was a nominal risk increase for melanoma and kidney cancer. There was no association between blood pressure and asthma, as expected, providing additional support for a lack of confounding. (text continues above)

Example 7.4. Use of a negative control exposure in a study of maternal alcohol consumption and hypertensive disorders of pregnancy

For the Avon Longitudinal Study of Parents and Children cohort, Martin et al. (2022) estimated the association of maternal alcohol consumption during pregnancy with hypertensive disorders of pregnancy. They used the mother's partner's alcohol consumption as an NCE and found that the alcohol intakes of both the mother and the mother's partner were associated with decreased odds of hypertensive disorders of pregnancy; this suggests that the findings were due to shared environmental exposures rather than a true causal effect of alcohol. (text continues on page 183)

In such studies, disease rates are typically estimated for particular industries or jobs and results are compared with rates for the general population (usually nationally). However, it is difficult to determine whether differences in cancer risk are due to the occupational exposure or due to differences in (unmeasured) lifestyle-related behaviours of the cohort participants.

One way to address this is to recruit internal reference groups, such as those working in the same plant but only in the office or those with short employment duration. If internal analyses are not possible, indirect methods can be used to evaluate the direction and magnitude of this unmeasured confounding. Steenland et al. (1984) outlined four simple methods that can be applied using readily available records. These are outlined in Example 7.5 for a study involving smelter workers and lung cancer.

A simple spreadsheet and code to help apply indirect control methods is available at https://sites.google.com/site/biasanalysis/Home (Fox et al., 2021).

(d) Application of g-methods to address time-varying confounding

Another issue that is often not adequately addressed is time-varying confounding. Although for many cohort studies data may be collected at multiple time points, researchers often use baseline measures of exposures and confounders to address causal questions pertaining to cancer risk. However, if the exposure changes over time, bias from inadequate adjustment for time-varying confounding may be problematic (Example 7.6).

7.4 Information bias due to exposure and outcome misclassification

This section first describes reporting considerations for study authors to report the data required to facilitate approaches to assess the direction and quantify the magnitude of measurement error and misclassification of exposure and outcome using only published data, as discussed in Chapter 4.

The second part of this section outlines the information that study authors can report to assist reviewers in determining the likelihood and magnitude of bias due to measurement error and exposure and outcome misclassification, and which biases should be prioritized in quantitative bias assessment.

The third part of this section briefly describes a selection of approaches that can be used to quantify information bias where access to individual-level study data is available, along with further resources about these approaches and examples of where they have been applied in studies of red meat consumption and mobile phone use.

7.4.1 Reporting considerations to facilitate information bias assessment

Chapter 4 describes a range of approaches that can be used with summary-level data to quantify bias caused by non-differential and differential error in the measurement of exposures and outcomes. Table 7.2 outlines the data that are needed to perform the bias assessment methods described in Chapter 4.

Note that almost all of the required information comes from validation studies. Such studies are important in providing the bias parameters that can be used to quantify bias with summary-level or individual-level data.

In addition to the reporting considerations outlined here, the authors of validation studies should also report their sampling, recruitment, and data collection methods, so that readers can assess the validity of the resulting bias parameters. The study authors should also report the characteristics of participants in any validation study, so that readers can assess the transportability of the bias parameters to other populations.

Example 7.5. Indirect methods to evaluate confounding in a cohort of lead smelter workers

First, if a cohort of lead smelter workers was found to have a higher risk of lung cancer than expected, one could examine whether the cohort also had an excess risk of other smoking-related diseases. If risk was elevated for the majority of smoking-related diseases (including diseases that are not thought to be affected by lead smelting), it is likely that the cohort smoked more than the general population did, and thus unmeasured confounding would explain the elevated risk of lung cancer. If the risk was not elevated for the majority of smoking-related diseases, smoking would be unlikely to be a strong confounder in the investigated exposure–outcome relation.

Second, rather than comparing the rate observed in the occupational cohort with that in the national population, another comparison group with a similar socioeconomic profile could be chosen. For example, one might expect that individuals working in lead smelters would have a similar socioeconomic position to workers in recycling plants. If the rates of lung cancer were similar between occupational cohorts of workers in smelters and recycling plants, this would suggest that smoking, rather than exposure to lead smelting, was increasing the risk of disease. However, this method is not appropriate if lead smelting causes the same cancers as smoking does, or if working in a recycling plant involved exposures to lung carcinogens. This alternative comparison of risk in a similar socioeconomic population is conceptually similar to using an NCE (Section 7.3.3(b)).

Third, adjustment can be made under different assumptions about the smoking behaviour of occupational cohort participants. Estimated rates of smoking in different occupational and sociodemographic groups are readily available. If there were a difference in smoking rates between lead smelter workers and the general population, one could adjust the risk estimate in a study of lung cancer accordingly. An illustration of such an indirect adjustment is given in Example 3.15.

Finally, another indirect method that can be used is to examine risk by years of exposure or by exposure levels. If working in a lead smelter increased the risk of developing lung cancer, one would expect to observe a dose–response effect by years of employment or exposure level. Ideally, this analysis should be stratified by age, so that workers within the same age categories are compared according to their duration of employment. One could assume that new workers would have been smoking for the same duration as long-term employees within the same age category. Thus, if no dose–response effect was noted, one might conclude that unmeasured confounding from smoking was present and that this explained the observed effect. One caution with this approach is that analyses based on measures of employment duration are particularly susceptible to healthy worker survivor bias (see Section 3.2.4(a) and Example 3.6). (text continues on page 184)

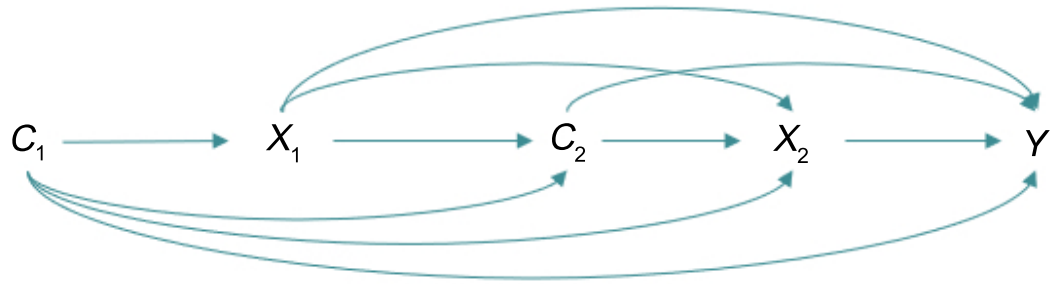
Table 7.2. Reporting considerations to facilitate bias assessment methods outlined in Section 4.3

Bias source	Bias parameters to report
Misclassification for a binary exposure	Sensitivity and specificity of the exposure measurement method, along with other potential bias parameters, such as positive and negative predictive values – by case or control status where applicable
Measurement error in continuous and categorized exposures	For studies that have used regression calibration, the attenuation factor and the validity coefficient (the correlation coefficient between the observed exposure and the true exposure) To facilitate the method of Rosner et al. (1990) (outlined in Section 4.3.5), when the exposure and confounders in the calibration equation and the exposure–outcome association are all linear, authors should report the coefficients of each variable in the calibration equation and each coefficient in the regression of outcome on the observed exposure and confounders.
All measurement error	To facilitate simple bias assessment methods (e.g. reallocation of counts of case and control participants), which are based on unadjusted results, authors should report both unadjusted and adjusted risk estimates.

Example 7.6. Bias from inadequate adjustment for time-varying confounding

In Fig. 7.3, red meat consumption is the time-varying exposure (X_1, X_2), body composition is the time-varying confounder (C_1, C_2), and colon cancer risk is the outcome Y. It is assumed that body composition (C_2) affects how much red meat someone eats (X_2), but it can be seen that this confounder (C_2) is also affected by prior exposure to red meat (X_1; exposure–confounder feedback). If body composition (C_2) is conditioned on, an intermediate variable on the causal pathway will have been adjusted for; this can produce biased estimates (Daniel et al., 2013). In contrast, if body composition (C_2) is not adjusted for, there is uncontrolled confounding. Conventional regression cannot adjust for time-varying confounding appropriately. Alternative statistical approaches, known as generalized methods (g-methods), are required to handle the issue of exposure–confounder feedback adequately (Robins and Hernán, 2009; Naimi et al., 2017). This is discussed further in Section 3.2.4(a).

Fig. 7.3. Directed acyclic graph demonstrating time-varying exposure in the presence of time-varying confounding. X_1, exposure at time 1; X_2, exposure at time 2; C_1, confounder at time 1; C_2, confounder at time 2; Y, outcome.



When confounders vary over time and are affected by prior exposure, they can also be mediators (see Example 2.1b). When researchers have access to individual-level data, there are opportunities to return to existing cohort studies and apply these methods to better answer causal questions, as in Example 7.7. (text continues on page 184)

Example 7.7. Use of g-methods to control for time-varying confounding in a study of titanium dioxide exposure

Bertke et al. (2021) reanalysed data from a cohort of 5163 boatbuilders exposed to styrene in Washington State in the USA who were employed between 1959 and 1978. Using g-estimation of a structural nested model to account for healthy worker survivor bias, they estimated that 1 year of exposure to styrene at a concentration of > 30 ppm accelerates time to lung cancer death by 2.3 years (95% confidence interval [CI], 1.53–2.94).

This analysis enabled estimation of the necessary components of the healthy worker survivor bias and provided evidence that this effect was potentially quite large, probably masking the true exposure–response relation in previous studies.

There are several reporting considerations for g-methods related to model specification. For example, researchers should compare the simulated risk of outcome under the natural course; a natural-course intervention is one that attempts to emulate the existing data by modelling the exposure in addition to confounders and outcomes. The results from the natural-course model can be compared with the observed risk as an informal validation of correct model specification. For detailed information, refer to Hernán and Robins (2020).

CAUSALab at the Harvard T.H. Chan School of Public Health maintains a repository of macros and code relevant to different g-methods (Harvard T.H. Chan School of Public Health, 2024). (text continues above)

7.4.2 Reporting considerations to facilitate evaluation of bias in individual studies

Table 7.3 summarizes reporting considerations that study authors can include in their manuscripts to assist reviewers and other readers in determining the likelihood and magnitude of bias from measurement error and misclassification of exposure and outcome, and which biases should be prioritized in quantitative bias assessment. Some of this information may also help to facilitate approaches (described in Chapter 4) that can be used to assess the direction and quantify the magnitude of measurement error and exposure and outcome misclassification using only published data. Further information about each of these biases can be found in Chapter 4.

Where the study authors believe that a particular form of bias is unlikely to have affected the observed results, the authors should provide an explanation for this assumption (e.g. Example 7.8).

7.4.3 Methods that can be used with individual-level data

Whereas the approaches outlined in Chapter 4 can be taken using summary-level data by the researchers themselves, by the study team analysing existing data, or by reviewers and hazard assessors, other approaches require access to individual-level data. Additional information beyond the primary study data may be required to quantify the effect of measurement error on estimated exposure–disease associations. Such data may come from internal validation studies conducted on a subset of the participants for whom (apparently) true exposure data are collected, or from external validation studies. Next, methods are briefly outlined that require individual-level data and that have previously been applied in studies used for hazard identification. Then, examples of studies that have used such approaches are briefly described.

Table 7.3. Reporting considerations for measurement error and exposure and outcome misclassification

Type of bias to be assessed	Reporting considerations	More details
Measurement error in binary exposures	Sensitivity and specificity of measures used to classify participants as exposed, along with relevant references	Section 4.2.1(b)
Measurement error in continuous exposures	Validity of exposure measurement, along with relevant references	Section 4.2.1(a)
Recall bias	Timing of measurement of exposure, in both case and control participants Exposure prevalence in general population	Section 4.2.3(a)
Interview or assessor error or bias	Interview quality by case or control status Whether methods used to assess or assign exposure status were blind to outcome status Whether case and control participants were assessed by the same interviewers or assessors Distribution of exposure across interviewers or assessors	Section 4.2.2
Proxy respondent bias	Percentage of proxy respondents in sample and in case and control participants Distribution of exposure in proxy and personal respondents	Section 4.2.3(b)
Reporting bias based on belief about a health hazard	Participants' beliefs about whether an exposure affects cancer risk, by case or control status where applicable	Section 4.2.3
Outcome misclassification	Source of all outcome data and sensitivity and specificity of the classification Whether methods used to assess or assign outcome status were blind to exposure status Where subtypes of specific cancers are analysed (e.g. specific types of non-Hodgkin lymphoma, or premenopausal and postmenopausal breast cancer), the basis on which subtypes were classified (e.g. specific International Classification of Diseases [ICD] version)	Section 4.4

Example 7.8. Explaining assumptions about differential sources of error

In a study on mobile phone use and the risk of brain tumours, Castaño-Vinyals et al. (2022) reported, "No formal analysis was conducted to take into account a possible differential recall bias, since the results of the operators' validation study provided no evidence for differential recall between [case and control participants]." (text continues on page 187)

(a) Classical non-differential exposure measurement error

Regression calibration is one of the more common approaches used to quantify and correct for classical measurement error with individual-level data. Regression calibration is described in detail in Section 4.3.6, along with situations where regression calibration approaches can be used to quantify measurement error using published data. To briefly recap, regression calibration involves using error-prone exposure variables (e.g. simple food frequency questionnaires to measure red meat consumption) and other participant characteristics that are available for the whole study population to predict the exposure obtained from a more accurate measurement (e.g. 24-hour diet recall) in a smaller sample. The calibration equation can then be applied to the whole study population and the resulting variable used as the exposure in the main analysis, with standard errors adjusted for the calibration. The resulting risk estimates can be compared with risk estimates from the original analysis (which used uncalibrated exposure variables) to assess the direction and magnitude of bias present (Example 7.9).

Further details about the implementation of regression calibration can be found in Fox et al. (2021, Chapter 10). Statistical software to conduct regression calibration is available in SAS (%blinplus macro) (Yale School of Public Health, 2024), Stata (merror package) (Stata, 2003), and R (merror package) (Bilonick, 2023).

As noted previously, other methods are available to quantify and correct for exposure measurement error, in addition to regression calibration. Some of these methods – simulation extrapolation for misclassification (MC-SIMEX), the Bayesian model

Example 7.9. Regression calibration to quantify bias due to measurement error

As noted in Example 4.22, regression calibration was used in the European Prospective Investigation into Cancer and Nutrition (EPIC), which was a cohort study. This example gives more detail on how this was done. Norat et al. (2005), in their investigation of consumption of red and processed meat and risk of colorectal cancer, quantified the impact of classical measurement error using individual-level data obtained from EPIC. In the EPIC study, all participants completed a self-administered dietary questionnaire, and an additional 24-hour diet recall measurement was taken from a random sample of 8% of the EPIC participants. Among the subsample, the 24-hour diet recall values for consumption of red and processed meat were regressed on the corresponding values obtained using the main dietary questionnaire, with a range of dietary and non-dietary factors included as covariates. Sex-specific and study-centre-specific calibration models were then applied to the whole cohort to predict values for the consumption of red and processed meat for each participant in the EPIC sample. These predicted values were then used in analyses to estimate the association between consumption of red and processed meat and colorectal cancer risk, with standard errors adjusted for the calibration; a stronger effect was observed after calibration (hazard ratio [HR], 1.55 for each 100 g increase; 95% CI, 1.19–2.02) than in the original analysis (HR, 1.25; 95% CI, 1.09–1.41). (text continues above)

for quantifying bias, and multiple imputation – are described briefly in Section 4.3.7; other methods are described in Keogh and White (2014).

(b) Differential measurement error

Section 4.3.5 outlines situations where probabilistic bias analysis can be used to quantify measurement error using published data. Probabilistic bias analysis can also be used to quantify differential (or non-differential) measurement error with individual-level data. The aim of probabilistic bias analysis is to provide bias-adjusted estimates over a plausible distribution of bias parameters, as opposed to a single value in simple bias analysis. The plausible distribution of bias parameters can be obtained from internal or external validation studies (Example 7.10).

Probabilistic bias analysis with individual-level data is covered in Chapter 9 of Fox et al. (2021). A range of software to conduct probabilistic bias analysis can be found at Columbia Mailman School of Public Health (2024).

Another approach that can be taken to evaluate the potential impact of differential measurement error, specifically recall bias in case–control studies, is the recruitment of different control groups in the analysis stage. This approach is described in Example 4.13. Briefly, this approach involves the recruitment of a control group for whom recall is likely to be similar to that of the case participants but who have a disease that is not thought to be associated with the exposure of interest (e.g. participants with a different cancer type). To undertake this analysis after data collection is complete would require the availability of information on exposure status for the new control group, as in Example 7.11.

Other methods that can be used with individual-level data to assess and quantify differential information bias include NCEs (see Section 7.3.3) and stratifying analyses by exposure causation belief, interviewer, or proxy respondent status. These methods are discussed further in Section 4.2.3.

7.5 Selection bias

Selection bias is a systematic error that might present a threat to a study's internal validity. Therefore, it is important that researchers carefully consider the potential for selection bias when analysing study data and identify and report the information necessary to assess the potential for such a bias, as well as its direction and magnitude. This could be included as part of the study results or as supplementary material.

Selection bias arises either by design or through analytical choice. As described in Chapter 5, cohort studies are prone to two main origins of selection bias. First, differential selection forces can drive a differential baseline participation or result in a differential loss to follow-up, so that results do not reflect the patterns in the source population. The second main origin of selection bias arises from left or right truncation during the analysis. These types of bias can also occur in case–control studies. In addition, bias can occur in case–control studies in the selection of control participants. For example, if the researchers

Example 7.10. Probabilistic bias analysis to quantify recall bias

Momoli et al. (2017) used case–control data from the Canadian part of the Interphone study to investigate mobile phone use and the risk of head and neck tumours. The main concern was recall bias regarding the use of mobile phones. Probability distributions for recall errors were derived from Interphone validation data, in which recalled mobile phone use was compared with operator records, separately for case and control participants (Vrijheid et al., 2006). A Monte Carlo procedure was then used to correct for recall bias, with the aim of recreating, as it were, the study population that would have been observed if recall bias were absent. A further sensitivity analysis was conducted to address possible bias with respect to the timing of interviews, because of concerns about this differing between case and control participants. The results of the probabilistic bias modelling were not meaningfully different from the results of the non-bias-adjusted analyses. (text continues above)

Example 7.11. Case–case analyses to quantify recall bias

Cardis et al. (2011) used a subset of data from the Interphone study to examine the associations between exposure to radiofrequency electromagnetic field (RF-EMF) radiation from mobile phone use and the risk of brain tumours. In that study, case–case analyses were conducted in which mobile phone use was compared between case participants with tumours of the brain in areas highly exposed to RF-EMF radiation and case participants with tumours in other parts of the brain with lower exposure. The case–case analysis showed increased odds ratios for tumours in the most exposed part of the brain in individuals with ≥ 10 years of mobile phone use (OR, 2.80; 95% CI, 1.13–6.94 for glioma), compared with other areas among long-term users, but no increased odds ratios for individuals who had started using a mobile phone more recently. (text continues on page 189)

select people with another disease as the control source population and if that disease is related to the exposure, then the control participants will not be representative of the source population for the case participants. This section outlines how researchers can examine selection biases in their own studies and facilitate the analysis of selection bias by reviewers who will not have access to the individual-level data.

7.5.1 Reporting considerations to facilitate assessment of selection bias by expert reviewers using methods outlined in Chapter 5

Table 7.4 summarizes the study information that should be reported to enable assessment of selection bias at a later stage, as described in Chapter 5.

7.5.2 Differential baseline participation

A simple bias analysis to address the effect of differential baseline participation (in both cohort and case–control studies) should be informed by internal data, reported as a contingency table of participation proportions for each of the combinations of exposure and disease. The prevalence or distribution of exposure and disease should also be estimated and reported for the non-participants (both case and control participants). Ideally, this estimate should be based on an internal validation study of a group of the non-participants. If such an internal validation study is not possible, it may be possible to provide estimates of the prevalence of exposure and disease in non-participants based on expert judgement or external data. If individual-level data are available, those external estimates can be applied to the study data (e.g. perhaps adjusting for age, sex, and other key subgroups of interest), as in Examples 7.12 and 7.13.

Internal validation substudies should be recognized as an important strength of study design. However, such substudies are not always possible. If the estimates of the exposure prevalence among non-participating case and control individuals cannot be informed by the internal data, external data or expert judgement can help in assigning values of selection proportions and conducting simple bias analysis. In this situation, it is important to report the sources and external data as well as the hypothesis or educated guesses used to quantify the exposure prevalence among correspondents, to enable calculation of the selection probability in each key subgroup. This strategy was successfully applied in the Interphone study (Vrijheid et al., 2009), where several combinations of selection probabilities were assigned under several hypothetical scenarios of mobile phone use among non-participants (Example 7.14).

A freely available spreadsheet (https://sites.google.com/site/biasanalysis/Home; Fox et al., 2021) is useful for easily calculating the bias-adjusted odds ratios using such bias parameters as exposure distributions and selection proportions, informed by internal data, by simulation, or by educated guesses. The spreadsheet used in Example 7.14, along with other available tools, is presented in detail in Lash et al. (2021), and the spreadsheet is provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat). It may be possible to extend this analysis by documenting exposure prevalence by each stratum of age and sex.

Table 7.4. Essential information that should be reported to inform assessment of selection bias

Origin of selection bias	What should be reported	More details
Differential baseline participation	Definitions and distributions of participants and non-participants among case and control groups Prevalence of exposure and disease for non-participants Probability of selection among each subgroup	Section 5.2.1
Loss to follow-up	Rates of loss to follow-up in key subgroups of interest by baseline exposure status	Section 5.2.2
Left truncation (prevalent exposures)	Time zero Proportions of study participants who were subject to prevalent exposures at baseline, and, ideally, how long these participants had been exposed for (minimum, median, maximum) before follow-up commenced	Sections 5.2.3, 5.2.4
Right truncation (insufficient follow-up)	Minimum, median, and maximum lengths of follow-up for study participants, from baseline, as well as corresponding times since first exposure	Section 5.2.5
Bias due to selection of control participants	Eligible control diseases and their distribution in the study sample Exposures of interest on which the choice of the control diseases was based Distribution of exposure prevalence in target population and other potential source populations	Section 5.3.3

7.5.3 Differential loss to follow-up

Differential loss to follow-up can be a second source of selection bias, because it arises from differences in continued study participation that are related to both the exposure and the health outcome. When the information on participants lost to follow-up is missing at random, the bias can be addressed using methods of multiple imputation. Otherwise, the information available about participants before their loss to follow-up can inform the bias analysis. To conduct a simple analysis of such a bias, one might apply either the outcome modelling method or inverse probability of attrition weighting. Both methods require knowledge of the number of participants lost to follow-up by exposure status to impute the information lost to follow-up from data available to researchers. Such data should, at a minimum, specify for each exposure stratum the total number of participants, the number of participants with an outcome of interest per exposure status, and the person-years (Example 7.15).

7.5.4 More-sophisticated methods to adjust for bias due to loss to follow-up in the original study

More-complex methods exist to adjust for selection bias and are frequently implemented by researchers. For example, in the DAG in Fig. 5.2, the unblocked backdoor path (X–V–U–Y) from the exposure X to the outcome Y could be blocked by adjusting for the observed covariate V in a standard regression model; this would eliminate selection bias due to loss to follow-up.

Another option is to use inverse probability of attrition weights (IPAWs), which have been increasingly used to adjust for bias due to loss to follow-up (Hernán et al., 2004; Weuve et al., 2012). The IPAW is specified as the inverse of the probability of remaining in the study, conditional on predictors of attrition. In Fig. 5.2, simple IPAWs could be generated as $1/\Pr(L = 0 \mid V = v)$, although in practice these weights will be conditioned on more predictors of loss and stabilized to reduce variance. The IPAWs are then used in a regression model of Y on X to produce an effect that is adjusted for loss to follow-up, without needing to include V in the model. A particular benefit of IPAW methods is that they can be used in situations where standard covariate control would fail. For example, conditioning the analysis on those not lost to follow-up ($L = 0$) induces a correlation between the exposure X and the unmeasured confounder U, which would bias the effect of X on Y. Attempting to adjust for V in a regression model would not remove the bias, because V is a collider along the path from X to U. By avoiding conditioning on V, IPAWs enable the removal of bias due to loss to follow-up in this situation (Hernán et al., 2004; Weuve et al., 2012).

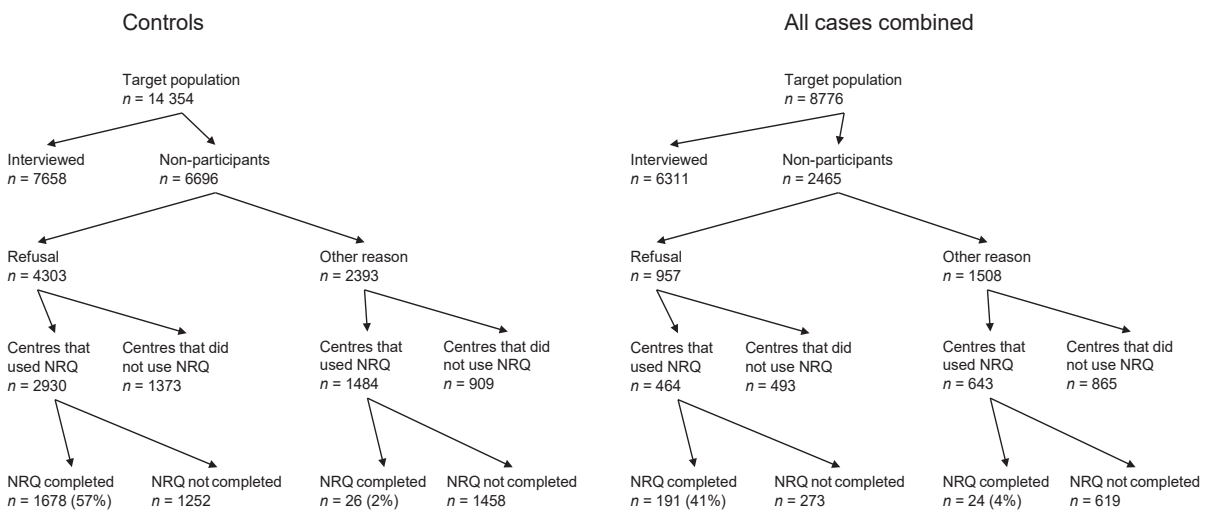
Example 7.12. Interphone study on mobile phone use and the risk of brain tumours

As discussed in Example 5.20, the multinational case–control Interphone study provides a good example of how to carefully examine the potential impact of selection bias. The aim of the study was to investigate whether RF-EMF radiation emitted by mobile phones increases the risk of brain tumours (Cardis et al., 2007). Eligible case participants were all residents of the study region, aged 30–59 years, who had been diagnosed during the study period with a first primary glioma, meningioma, or acoustic neuroma, confirmed either histologically or using unequivocal diagnostic imaging. Control participants were selected randomly from the same source population as case participants and matched to them by age, sex, and region.

The authors provided a comprehensive description of the study population with precise definitions of the study regions and the sizes of the source populations of case and control participants for each of 16 study regions in 13 participating countries (Table 1 of Cardis et al., 2007). Moreover, being aware that selection bias is a concern when inclusion is conditioned on consent to participate, the authors asked those who declined to participate to complete a short non-response questionnaire (NRQ), to estimate the prevalence of mobile phone use among non-participants (Vrijheid et al., 2009). The question about regular use of mobile phones on the NRQ was phrased as, "Have you ever used a mobile phone regularly? Yes or no?" Regular use was defined as use at least once a week for a period of 6 months or longer.

The authors provided detailed tables with definitions and distributions of participants and non-participants among case and control groups in the Interphone study (Table 2 of Vrijheid et al., 2009), along with the percentage distribution of regular mobile phone users among interviewed subjects (i.e. participants) and NRQ respondents (Table 3 of Vrijheid et al., 2009). Moreover, a flowchart of enrolment in the Interphone study given in an appendix (reproduced in Fig. 7.4), which reported participation frequencies for the case and control groups, facilitated calculation of the fraction of individuals in each category (i.e. interviewed participants, refusal with NRQ, refusal without NRQ, and other non-participants, as untraceable, ill, deceased, or other reason). This is important when estimating the probability of selection among those who do and do not use mobile phones (Table 3 of Vrijheid et al., 2009).

Fig. 7.4. Flow of subject enrolment into Interphone study. NRQ, non-response questionnaire. Source: Reprinted from Vrijheid et al. (2009), Copyright 2009, with permission from Elsevier.



Example 7.12. Interphone study on mobile phone use and the risk of brain tumours (continued)

Based on the reported distributions from Fig. 7.4 and the information that regular mobile phone use was reported by 69% of interviewed control participants, 56% of NRQ control participants, 66% of interviewed case participants, and 50% of NRQ case participants, one can produce a contingency table showing the participation and mobile phone use among case and control participants (Table 7.5).

Table 7.5. Participation and mobile phone use in the Interphone study[a]

	Participants	Non-participants with NRQ		Non-participants without NRQ	
	Regular use	No use	Regular use	No use	Cannot categorize
Case participants	2616	1348	105	105	2250
Control participants	3758	1688	951	748	4992

NRQ, non-response questionnaire.

[a] All types of brain tumour (i.e. glioma, meningioma, or acoustic neuroma) are combined. Numbers of non-participants with NRQ include both refusers and other non-participants.

Source: Observed aggregated data from Vrijheid et al. (2009).

From the data in Table 7.5, one can see that the odds of participation depend on disease status; the odds ratio is calculated as

$$\mathrm{OR} = \left(\frac{2616 + 1348}{105 + 105}\right) \Big/ \left(\frac{3758 + 1688}{951 + 748}\right) = 5.88 \qquad \text{(E7.1)}$$

meaning that the chance of participation in the case group is 5.88 times that in the control group. Participation also depends on exposure status, although to a lesser extent, with

$$\mathrm{OR} = (3758/951)/(1688/748) = 1.75 \qquad \text{(E7.2)}$$

It is noteworthy that this exposure status odds ratio is examined in control participants only.

The unadjusted odds ratio associating regular mobile phone use with brain tumour occurrence among study participants is

$$\mathrm{OR}_{\text{participants}} = (2616/3758)/(1348/1688) = 0.87 \qquad \text{(E7.3)}$$

This odds ratio is quite similar to the matched odds ratios observed for the original national and combined studies (Lahkola et al., 2007, 2008; Schoemaker et al., 2005).

Among non-participants who completed the NRQ, the unadjusted odds ratio is

$$\mathrm{OR}_{\text{non-participants}} = (105/951)/(105/748) = 0.79 \qquad \text{(E7.4)}$$

which is in the same direction as, but smaller than, the unadjusted odds ratio observed among participants. Consequently, the potential impact of selection bias seems to be rather limited in this example.

To verify this, one might further estimate the bias-adjusted odds ratio, by assuming that non-participants who did not complete the NRQ had the same exposure prevalence, conditional on case or control status, as those who completed the NRQ. To accomplish this solution, the numbers of non-participants who did not complete the NRQ in Table 7.5 were weighted using the exposure prevalence of the non-participants who completed the NRQ (Table 7.6).

Example 7.12. Interphone study on mobile phone use and the risk of brain tumours (continued)

Table 7.6. Participation and mobile phone use in the Interphone study with data from NRQ respondents projected to participants without NRQ[a]

Disease or exposure	Participants		Non-participants with NRQ		Non-participants without NRQ	
	Regular use	No use	Regular use	No use	Projected regular use	No use
Case participants	2616	1348	105	105	1125	1125
Control participants	3758	1688	951	748	2796	2196

NRQ, non-response questionnaire.

[a] All types of brain tumour (i.e. glioma, meningioma, or acoustic neuroma) are combined. Numbers of non-participants with NRQ include both refusers and other non-participants.

Source: Observed aggregated data from Vrijheid et al. (2009).

Data from Table 7.6 enable relatively easy estimation of the bias-adjusted odds ratio (OR, 0.92) and its comparison with the unadjusted odds ratio among full participants (OR, 0.87). Such a comparison would enable reviewers to conclude that the differential selection had not had a substantial effect on the estimated association between regular mobile phone use and brain tumour occurrence in this example. In fact, the odds ratio is slightly closer to the null; when confidence intervals are calculated, there could be weaker evidence for an association if all eligible individuals have been taken into account. (text continues on page 190)

Example 7.13. Mobile phone use and the risk of uveal melanoma

In this study – in which exposure prevalence was also assessed and reported using the NRQ, but only among non-participant control individuals – it was possible to identify a substantial bias due to selective participant selection (Lash et al., 2021). Regular mobile phone use was more prevalent among participating control individuals (45% in men and 25% in women) than among non-participating control individuals (37% in men and 16% in women) (Stang et al., 2009, Supplementary Table 3). The unadjusted odds ratio for association of regular mobile phone use with uveal melanoma was 0.71 among all participants and 1.26 among non-participants who completed the NRQ (Lash et al., 2021). The bias-adjusted odds ratio was estimated to be 1.62, suggesting that differential selection could have had a substantial impact on the effect estimate in this study by biasing it downwards. (text continues on page 190)

Example 7.14. Selection probabilities in the Interphone study

In this study, the authors reported several combinations of selection probabilities, which were assigned under several hypothetical scenarios of mobile phone use among non-participants (Table 7.7).

Table 7.7. Hypothetical scenarios of regular mobile phone use among non-participants in the Interphone study, as a function of observed use patterns in interviewed participants and NRQ respondents: glioma study

Scenario		Observed phone use (%)		Assumed phone use (basis for assumption) (%)		Assumed phone use in target population (%)	Selection probability	
		Inter-viewed	Refusal with NRQ	Refusal without NRQ	Other non-participants			
		P_1	P_2	P_3	P_4	P_{1-4}	S_1	S_0
Control participants	Fraction of subjects in each category W_1–W_4	0.53	0.17[a]	0.13	0.17	1.00		
R	Reference	69	69 (P_1)	69 (P_1)	69 (P_1)	69	0.53	0.53
A	NRQ applies to refusers with NRQ, unbiased use in other non-participants	69	56	66 [$m_w(P_{1-2})$]	66 [$m_w(P_{1-2})$]	66	0.55	0.48
B	NRQ applies to all refusers, unbiased use in other non-participants	69	56	56 (P_2)	64 [$m_w(P_{1-3})$]	64	0.57	0.46
C	NRQ applies to refusers with NRQ, 33% less use in other non-participants	69	56	46 (0.67 × P_1)	46 (0.67 × P_1)	60	0.61	0.41
D	NRQ applies to refusers with NRQ, 20% more use in other non-participants	69	56	83 (1.2 × P_1)	83 (1.2 × P_1)	71	0.52	0.57
E	NRQ applies to all non-participants	69	56	56 (P_2)	56 (P_2)	63	0.58	0.44
Cases of glioma	Fraction of subjects in each category W_1–W_4	0.64	0.05[a]	0.06	0.24	1.00		
r	Reference	65	65 (P_1)	65 (P_1)	65 (P_1)	65	0.64	0.64
a	NRQ applies to refusers with NRQ, unbiased use in other non-participants	65	53	64 [$m_w(P_{1-2})$]	64 [$m_w(P_{1-2})$]	64	0.65	0.63

Example 7.14. Selection probabilities in the Interphone study (continued)

Table 7.7. Hypothetical scenarios of regular mobile phone use among non-participants in the Interphone study, as a function of observed use patterns in interviewed participants and NRQ respondents: glioma study (continued)

Scenario		Observed phone use (%)		Assumed phone use (basis for assumption) (%)		Assumed phone use in target population (%)	Selection probability	
		Inter-viewed	Refusal with NRQ	Refusal without NRQ	Other non-participants			
		P_1	P_2	P_3	P_4	P_{1-4}	S_1	S_0
b	NRQ applies to all refusers, unbiased use in other non-participants	65	53	53 (P_2)	63 [$m_w(P_{1-3})$]	63	0.66	0.61
c	NRQ applies to refusers with NRQ, 33% less use in other non-participants	65	53	43 (0.67 × P_1)	43 (0.67 × P_1)	58	0.59	0.44
d	NRQ applies to refusers with NRQ, 20% more use in other non-participants	65	53	78 (1.2 × P_1)	78 (1.2 × P_1)	69	0.50	0.60
e	NRQ applies to all non-participants	65	53	53 (P_2)	53 (P_2)	61	0.69	0.58

NRQ, non-response questionnaire; m_w, weighted mean; P_1, prevalence of mobile phone use among interviewed subjects; P_2, prevalence of mobile phone use among refusers who completed the NRQ (P_2 = 0.82 × P_1: NRQ results for all case and control participants combined. The P_2/P_1 ratio was assumed to be the same for control and case participants and for different study centres and sex and age categories because the NRQ results did not indicate substantial or consistent differences between these groups. Data analysed from study centres with NRQ data were applied to all centres.); P_3, prevalence of mobile phone use among refusers who did not complete the NRQ; P_4, prevalence of mobile phone use among subjects who did not participate for a reason other than refusal (dead, too ill, physician refusal, untraceable, other); S_1, probability of selection (i.e. participation in full interview) among mobile phone users, $(W_1 \times P_1)/P_{1-4}$; S_0, probability of selection (i.e. participation in full interview) among non-mobile phone users, $[W_1 \times (1 - P_1)]/(1 - P_{1-4})$; W_1–W_4, fraction of total number of subjects ascertained in each response category for all study centres combined.

[a] W_2 is based on the fraction of NRQs completed for refusers in study centres that used the NRQ (57% in control participants, 41% in case participants).

Source: Reproduced from Vrijheid et al. (2009).

For instance, scenario C, for which it was assumed that other non-participants had a 33% lower prevalence of mobile phone use than interviewed subjects, was informed by external data, based on a comparison in Finland of the percentage of interviewed subjects and non-participants who had listed mobile phone numbers (Lahkola et al., 2005). Scenario D, for which it was assumed that other non-participants had a 20% higher prevalence of mobile phone use than interviewed subjects, was an educated guess (Vrijheid et al., 2009).

The reported data and selection probability make it easy to estimate a bias factor for each scenario using the formula proposed by Greenland and Criqui (1981). (text continues on page 190)

Example 7.15. Loss to follow-up and the association between shift work and breast cancer

The Nurses' Health Study was initially established in 1976. In the Nurses' Health Study II (NHS2, 1989–2013), 114 559 nurses completed the original questionnaire on shift work (Wegrzyn et al., 2017) to provide updated values on shift work. In the highest category of years of night shift work, drawing on the updated shift work history, those who had been followed up for ≤ 10 years had a multivariable-adjusted hazard ratio of 2.13 (95% CI, 1.19–3.81) and those with > 10 years of follow-up had a hazard ratio of 1.19 (95% CI, 0.78–1.81). Given that dropping out of the study is associated with outcome, a quantitative bias analysis of these data would be useful.

For this analysis, it is necessary to know the total number of participants who dropped out, the exposure status of those who dropped out, and the person-years of follow-up. The number of participants who had dropped out and their exposure status was not given; however, only about half of the total person-years (1 213 546/2 190 678 = 55%) were accumulated in those who were followed up for > 10 years, as shown in Table 7.8, which is excerpted from Table 3 of Wegrzyn et al. (2017). This implies that a considerable proportion of the original participants dropped out.

Table 7.8. Multivariable-adjusted associations between updated duration of rotating night shift work and invasive breast cancer, stratified by follow-up period, in the Nurses' Health Study II, 1989–2013

Exposure measure: cumulative years if rotating (updated) shift work	No. of case participants	No. of person-years	Age-adjusted		Multivariable-adjusted[a]		
			HR	95% CI	HR	95% CI	*P* for trend
≤ 10 years of follow-up							
None	341	321 600	1.00	Referent	1.00	Referent	
1–9	621	602 095	0.98	0.86–1.12	0.97	0.85–1.11	
10–19	60	50 481	0.92	0.70–1.21	0.94	0.71–1.23	
≥ 20	12	2 956	1.99	1.11–3.56	2.13	1.19–3.81	
All subjects (NHS2 cumulative rotating night shift work, updated), years[b]	1034	977 132					0.75
> 10 years of follow-up							
None	609	346 804	1.00	Referent	1.00	Referent	
1–9	1381	767 303	1.06	0.96–1.16	1.07	0.97–1.18	
10–19	141	88 801	0.90	0.74–1.07	0.95	0.79–1.14	
≥ 20	23	10 637	1.10	0.72–1.66	1.19	0.78–1.81	
All subjects (NHS2 cumulative rotating night shift work, updated), years[b]	2154	1 213 546					

CI, confidence interval; HR, hazard ratio; NHS2, Nurses' Health Study II.

[a] Multivariable-adjusted models were adjusted for the following covariates: age, height, body mass index, body mass index at age 18 years, adolescent body size, age at menarche, age at first birth and parity combined, breastfeeding, type of menopause and age at menopause combined, menopausal hormone therapy use, duration of use of menopausal hormonal therapy with estrogen alone, duration of use of estrogen and progesterone menopausal hormone therapy, first-degree family history of breast cancer, history of benign breast diseases, alcohol consumption, physical activity level, and current mammography. All categorical covariates were included in models with missing indicators.

[b] Analyses using updated data on duration of shift work excluded participants during the cycles in which they were missing information on shift work exposure, resulting in fewer case participants and person-years than in analyses using history of shift work reported at baseline in 1989. Values do not sum to the total because of rounding.

Source: Excerpted from Wegrzyn et al. (2017).

Example 7.15. Loss to follow-up and the association between shift work and breast cancer (continued)

Therefore, it is possible to calculate a crudely adjusted result for each stratum. This is done by reweighting the person-time to account for a presumed continuation of the risk in those lost to follow-up.

Table 7.9 shows the calculations for those with ≥ 10 years of shift work. It is assumed that the total number of person-years and of cancers in those lost to follow-up are twice the number seen (i.e. the risk stayed the same in the years after the 10 years of follow-up). Then the imputed total number of subjects who had been followed up for > 10 years consists of the sum of the number with complete follow-up plus twice the number lost to follow-up. The resulting crude hazard ratio is 2.18, which is higher than that calculated for the group with complete follow-up (HR, 1.23). This suggests that loss to follow-up has downwardly biased the hazard ratio that would have been observed if there were no loss to follow-up. (text continues on page 191)

Table 7.9. Imputation of hazard ratios to account for loss to follow-up in the Nurses' Health Study II

	Complete follow-up > 10 years of follow-up		Lost to follow-up ≤ 10 years of follow-up		Imputed total 2 × lost to follow-up + complete follow-up	
Shift work	None	≥ 20	None	≥ 20	None	≥ 20
Breast cancers	609	23	341	12	1291	47
People (assume half of original cohort dropped out)	21 764.5	81	21 764.5	81	43 529	162
Person-years	346 804	10 637	321 600	2956	990 004	16 549
Crude rate per 100 000 person-years	1756	2162	106	406	130.4	284
Crude rate difference		40.6		299.9		153.6
Crude rate ratio		1.23		3.83		2.18

7.5.5 Bias due to selection of control participants in case–control studies

In a case–control study, bias can arise if the control and case participants are chosen from different source populations. This section outlines how researchers can assess the direction and magnitude of bias in the selection of control participants, when hospitalized patients are recruited as control participants. A full explanation of the rationale and methods can be found in Section 5.3.2 and is summarized here. The example given is that of the recruitment of hospital control participants, but it is important to understand that the same hypothetical selection biases may occur for other sources of control participants.

The ideal case–control study recruits control participants from the same source population as the case participants. The source population is not always easy to define or to access, so in some situations, researchers recruit hospital patients as control participants. In these situations, two selection phases have occurred: (i) the selection of control participants from the source population into the hospital, and (ii) their selection from the hospital into the study group.

The selection into the hospital could be affected by a wide range of factors. Socioeconomic status may affect who enters the hospital, particularly for less-severe conditions, treatments that are optional (e.g. some plastic surgery), or treatments that can be performed either as day procedures or with hospital admission. The area served by the hospital may differ according to the disease; for example, if a hospital specializes in treating a particular cancer, the source population for people with that cancer may come from a wider geographical area than for people hospitalized for non-cancer reasons. In addition, hospital patients are more likely to have exposures that lead to the disease they are hospitalized for, as

well as leading to the case disease. In this type of study, it is important to report the proportion of participants with the exposure (by age, sex, or other relevant variables), so that a comparison can be made with other data. Example 7.16 examines potential bias arising from the recruitment of hospital control participants.

More-sophisticated adjustments can be made by adjusting for the prevalence of the exposure within subgroups of the population. For example, if researchers were interested in differences between men and women and a previous survey had published rates of opium exposure by age and sex subgroups, a stratified analysis could be performed.

The same approach can be used for other situations when different selection factors are operational in the selection of the control and case participants, for example if friends are recruited as control participants or there are different (and biased) participation fractions in the case and control participants.

7.6 Conclusions

This chapter is aimed at researchers who have access to individual-level data and wish to undertake a quantitative bias assessment. It follows the order of the previous chapters in this volume, covering, in turn, the use of graphical tools to assess bias and methods to quantitatively assess confounding, information bias (measurement error and misclassification), and selection bias. For each of these sections, methods mentioned in the previous chapters are identified that could be used to undertake a

Example 7.16. Case–control study of opium exposure and oesophageal cancer

In a study by Shakeri et al. (2012), also described in Examples 4.14, 5.18, 5.22, 5.27, and 5.29, control participants were selected from the same hospital as the case participants and were individually matched on age and sex. Control participants were selected from those inpatients with diseases thought to be unrelated to tobacco use, alcohol consumption, or diet, because these factors were thought to be related to oesophageal cancer. The question to be addressed is whether opium exposure is more likely in the hospital-based control participants than in the neighbourhood from which the case participants arose. If so, it is necessary to determine the magnitude and direction of the resultant bias in the study.

As an initial simple analysis, the prevalence of opium smoking in the neighbourhood can be used to calculate the expected distribution of opium exposure in the control participants (Table 7.10). The spreadsheet used in this example is provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat). The number of unexposed and exposed control participants can be weighted by the prevalence of opium smoking in the neighbourhood. This adjustment results in an odds ratio of 2.41, compared with the original unadjusted odds ratio of 1.36. This suggests that the recruitment of hospital control participants markedly biased the association towards the null. (text continues above)

Table 7.10. Bias adjustment of odds ratios calculated for hospital-based control participants by applying neighbourhood exposure prevalence

	Hospitalized case participants	Hospitalized control participants	Odds ratio
Opium smokers	45	73	1.36
Non-opium smokers	85	187	
Hospital prevalence of opium smoking (%)	35	28	
Neighbourhood prevalence of opium smoking (%)		18	
Opium smokers (expected)	45 (no change)	46.8	2.41
Non-opium smokers (expected)	85 (no change)	213.2	

Source: Lash et al. (2009, p. 51).

quantitative bias analysis by researchers who have access to individual data. In addition, types of data that should be reported to facilitate bias assessment in future systematic reviews and hazard identification documents are recommended. Finally, statistical packages, spreadsheets, and code that are available to help researchers undertake quantitative bias assessments are suggested.

It is hoped that this chapter will assist researchers in undertaking quantitative bias assessments in their own studies. It is also anticipated that epidemiologists will increasingly return to existing large cohort studies to apply newer conceptual and statistical methods to address causal questions pertaining to cancer risk and survival. The inclusion of quantitative bias assessment should be an integral component of every epidemiological study. It is hoped that the information provided in this chapter will assist researchers in determining the magnitude and direction of bias in all their studies, and that the reporting of the factors needed to undertake such analyses will facilitate stronger systematic reviews and hazard identifications.

References

Axelson O, Steenland K (1988). Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med. 13(1):105–18. doi:10.1002/ajim.4700130107 PMID:3344750

Bertke SJ, Keil AP, Daniels RD (2021). Lung cancer mortality and styrene exposure in the reinforced-plastics boatbuilding industry: evaluation of healthy worker survivor bias. Am J Epidemiol. 190(9):1784–92. doi:10.1093/aje/kwab108 PMID:33847736

Bilonick RA (2023). merror: accuracy and precision of measurements, version 3.0. Available from: https://cran.r-project.org/web/packages/merror/index.html.

Cardis E, Armstrong BK, Bowman JD, Giles GG, Hours M, Krewski D, et al. (2011). Risk of brain tumours in relation to estimated RF dose from mobile phones: results from five Interphone countries. Occup Environ Med. 68(9):631–40. doi:10.1136/oemed-2011-100155 PMID:21659469

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

Castaño-Vinyals G, Sadetzki S, Vermeulen R, Momoli F, Kundi M, Merletti F, et al. (2022). Wireless phone use in childhood and adolescence and neuroepithelial brain tumours: results from the international MOBI-Kids study. Environ Int. 160:107069. doi:10.1016/j.envint.2021.107069 PMID:34974237

Chan II, Kwok MK, Schooling CM (2021). Blood pressure and risk of cancer: a Mendelian randomization study. BMC Cancer. 21(1):1338. doi:10.1186/s12885-021-09067-x PMID:34915881

Columbia Mailman School of Public Health (2024). Probabilistic sensitivity analysis of misclassification. New York (NY), USA: Columbia Mailman School of Public Health. Available from: https://www.publichealth.columbia.edu/research/population-health-methods/probabilistic-sensitivity-analysis-misclassification.

Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA (2013). Methods for dealing with time-dependent confounding. Stat Med. 32(9):1584–618. doi:10.1002/sim.5686 PMID:23208861

Diallo A, Deschasaux M, Latino-Martel P, Hercberg S, Galan P, Fassier P, et al. (2018). Red and processed meat intake and cancer risk: results from the prospective NutriNet-Santé cohort study. Int J Cancer. 142(2):230–7. doi:10.1002/ijc.31046 PMID:28913916

English DR, MacInnis RJ, Hodge AM, Hopper JL, Haydon AM, Giles GG (2004). Red meat, chicken, and fish consumption and risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev. 13(9):1509–14. doi:10.1158/1055-9965.1509.13.9 PMID:15342453

Fox MP, MacLehose R, 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

Greenland S, Criqui MH (1981). Are case-control studies more vulnerable to response bias? Am J Epidemiol. 114(2):175–7. doi:10.1093/oxfordjournals.aje.a113180 PMID:7304552

Greenland S, Pearce N (2015). Statistical foundations for model-based adjustments. Annu Rev Public Health. 36(1):89–108. doi:10.1146/annurev-publhealth-031914-122559 PMID:25785886

Harvard T.H. Chan School of Public Health (2024). CAUSALab. Software. Boston (MA), USA: Harvard T.H. Chan School of Public Health. Available from: https://causalab.sph.harvard.edu/software/.

Hernán MA (2016). Does water kill? A call for less casual causal inferences. Ann Epidemiol. 26(10):674–80. doi:10.1016/j.annepidem.2016.08.016 PMID:27641316

Hernán MA, Hernández-Díaz S, Robins JM (2004). A structural approach to selection bias. Epidemiology. 15(5):615–25. doi:10.1097/01.ede.0000135174.63482.43 PMID:15308962

Hernán MA, Robins JM (2006). Instruments for causal inference: an epidemiologist's dream? Epidemiology. 17(4):360–72. doi:10.1097/01.ede.0000222409.00878.37 PMID:16755261

Hernán MA, Robins JM (2020). Causal inference: what if. Boca Raton (FL), USA: Chapman and Hall/CRC. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

Keogh RH, White IR (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology. Stat Med. 33(12):2137–55. doi:10.1002/sim.6095 PMID:24497385

Lahkola A, Auvinen A, Raitanen J, Schoemaker MJ, Christensen HC, Feychting M, et al. (2007). Mobile phone use and risk of glioma in 5 North European countries. Int J Cancer. 120(8):1769–75. doi:10.1002/ijc.22503 PMID:17230523

Lahkola A, Salminen T, Auvinen A (2005). Selection bias due to differential participation in a case-control study of mobile phone use and brain tumors. Ann Epidemiol. 15(5):321–5. doi:10.1016/j.annepidem.2004.12.009 PMID:15840544

Lahkola A, Salminen T, Raitanen J, Heinävaara S, Schoemaker MJ, Christensen HC, et al. (2008). Meningioma and mobile phone use – a collaborative case-control study in five North European countries. Int J Epidemiol. 37(6):1304–13. doi:10.1093/ije/dyn155 PMID:18676984

Lash TL, Fox MP, Fink AK (2009). Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 1st ed. New York (NY), USA: Springer. Available from: doi:10.1007/978-0-387-87959-8

Lash TL, Fox MP, Fink AK (2021). Applying quantitative bias analysis to epidemiological data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer Nature.

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

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

Martin FZ, Fraser A, Zuccolo L (2022). Alcohol intake and hypertensive disorders of pregnancy: a negative control analysis in the ALSPAC cohort. J Am Heart Assoc. 11(19)e025102. doi:10.1161/JAHA.121.025102 PMID:36172928

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

Moreno-Betancur M (2021). The target trial: a powerful device beyond well-defined interventions. Epidemiology. 32(2):291–4. doi:10.1097/EDE.0000000000001318 PMID:33323746

Naimi AI, Cole SR, Kennedy EH (2017). An introduction to g methods. Int J Epidemiol. 46(2):756–62. PMID:28039382

Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, Mazuir M, et al. (2005). Meat, fish, and colorectal cancer risk: the European Prospective Investigation into Cancer and Nutrition. J Natl Cancer Inst. 97(12):906–16. doi:10.1093/jnci/dji164 PMID:15956652

Robins J, Hernán M (2009). Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, editors. Advances in longitudinal data analysis. Boca Raton (FL), USA: Chapman and Hall; pp. 553–99.

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

Savitz DA, Wellenius GA, Trikalinos TA (2019). The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. Am J Epidemiol. 188(9):1581–5. doi:10.1093/aje/kwz131 PMID:31145434

Schoemaker MJ, Swerdlow AJ, Ahlbom A, Auvinen A, Blaasaas KG, Cardis E, et al. (2005). Mobile phone use and risk of acoustic neuroma: results of the Interphone case-control study in five North European countries. Br J Cancer. 93(7):842–8. doi:10.1038/sj.bjc.6602764 PMID:16136046

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

Shiba K, Kawahara T (2021). Using propensity scores for causal inference: pitfalls and tips. J Epidemiol. 31(8):457–63. doi:10.2188/jea.JE20210145 PMID:34121051

Stang A, Schmidt-Pokrzywniak A, Lash TL, Lommatzsch PK, Taubert G, Bornfeld N, et al. (2009). Mobile phone use and risk of uveal melanoma: results of the risk factors for uveal melanoma case-control study. J Natl Cancer Inst. 101(2):120–3. doi:10.1093/jnci/djn441 PMID:19141780

Stata (2003). Stata software for generalized linear measurement error models. Available from: https://www.stata.com/merror/.

Steenland K, Beaumont J, Halperin W (1984). Methods of control for smoking in occupational cohort mortality studies. Scand J Work Environ Health. 10(3):143–9. doi:10.5271/sjweh.2349 PMID:6474108

Steenland K, Schubauer-Berigan MK, Vermeulen R, Lunn RM, Straif K, Zahm S, et al. (2020). Risk of bias assessments for evidence syntheses of observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. Environ Health Perspect. 128(9):95002. doi:10.1289/EHP6980 PMID:32924579

Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. (2021). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol. 50(2):620–32. doi:10.1093/ije/dyaa213 PMID:33330936

Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GTH (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 45(6):1887–94. Available from https://dagitty.net. PMID:28089956

VanderWeele TJ (2019). Principles of confounder selection. Eur J Epidemiol. 34(3):211–9. doi:10.1007/s10654-019-00494-6 PMID:30840181

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

Vrijheid M, Richardson L, Armstrong BK, Auvinen A, Berg G, Carroll M, et al. (2009). Quantifying the impact of selection bias caused by nonparticipation in a case-control study of mobile phone use. Ann Epidemiol. 19(1):33–41. doi:10.1016/j.annepidem.2008.10.006 PMID:19064187

Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, et al. (2019). A systematic review: tools for assessing methodological quality of human observational studies. Canberra, Australia: National Health and Medical Research Council. Available from: https://nhmrc.gov.au/guidelinesforguidelines/develop/assessing-risk-bias.

Wegrzyn LR, Tamimi RM, Rosner BA, Brown SB, Stevens RG, Eliassen AH, et al. (2017). Rotating night-shift work and the risk of breast cancer in the Nurses' Health Studies. Am J Epidemiol. 186(5):532–40. doi:10.1093/aje/kwx140 PMID:28541391

Weisskopf MG, Webster TF (2017). Trade-offs of personal versus more proxy exposure measures in environmental epidemiology. Epidemiology. 28(5):635–43. doi:10.1097/EDE.0000000000000686 PMID:28520644

Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, et al. (2012). Accounting for bias due to selective attrition: the example of smoking and cognitive decline. Epidemiology. 23(1):119–28. doi:10.1097/EDE.0b013e318230e861 PMID:21989136

Yale School of Public Health (2024). Center for Methods in Implementation and Prevention Science (CMIPS). Available from: https://ysph.yale.edu/cmips/research/software/blinplus/.