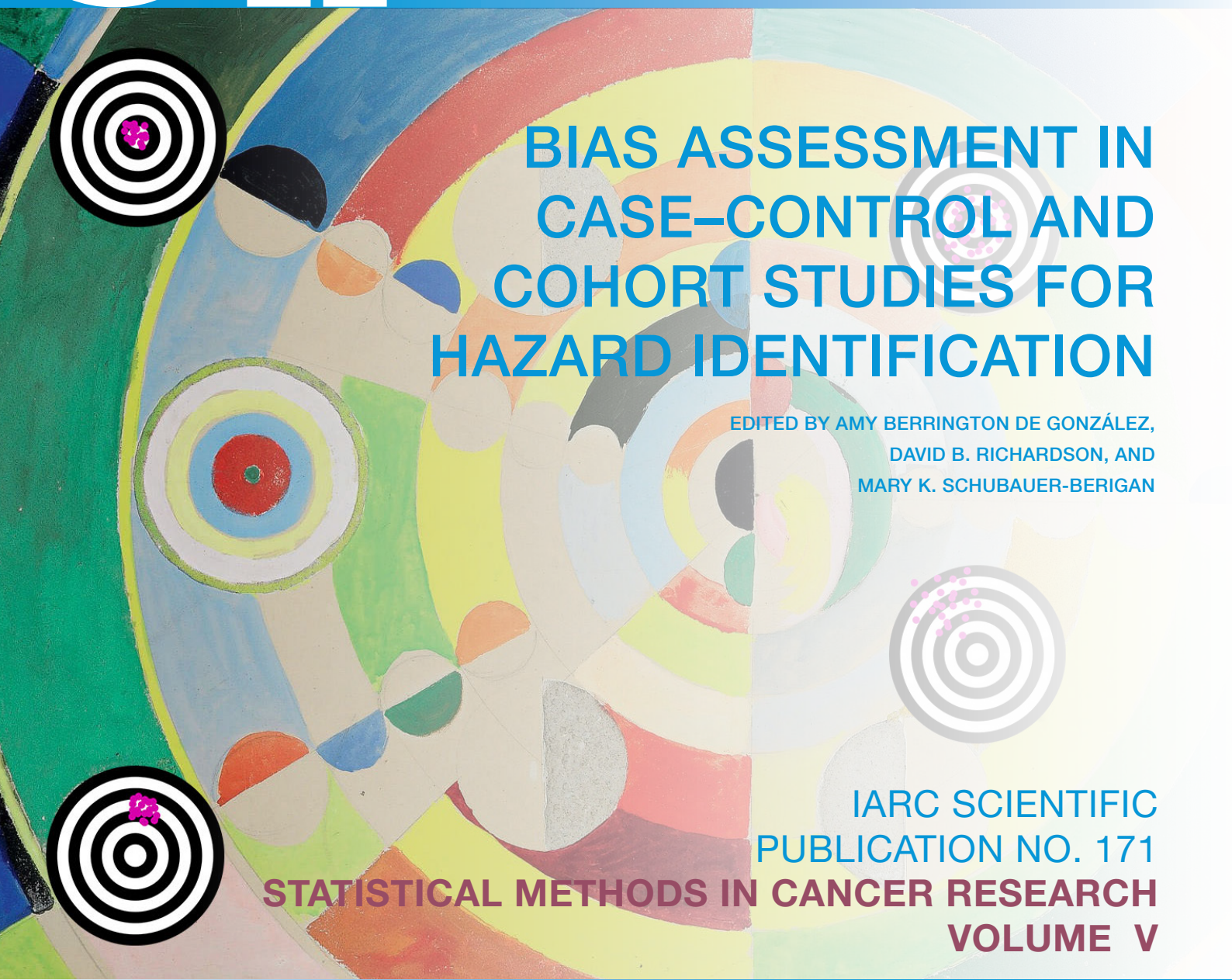
IARC SCIENTIFIC PUBLICATIONS



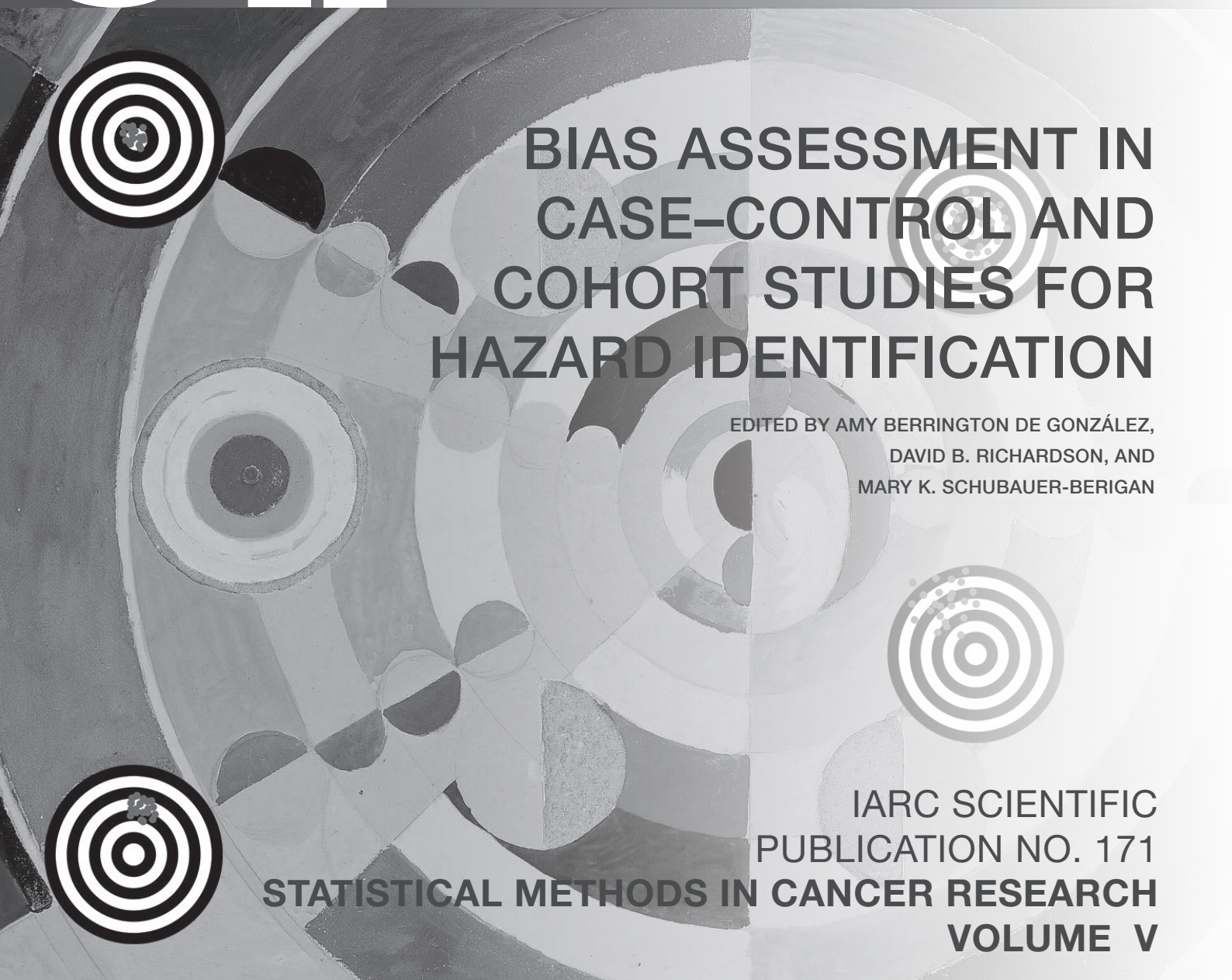
BIAS ASSESSMENT IN CASE–CONTROL AND COHORT STUDIES FOR HAZARD IDENTIFICATION

EDITED BY AMY BERRINGTON DE GONZÁLEZ,
DAVID B. RICHARDSON, AND
MARY K. SCHUBAUER-BERIGAN

IARC SCIENTIFIC
PUBLICATION NO. 171
STATISTICAL METHODS IN CANCER RESEARCH
VOLUME V

International Agency for Research on Cancer
World Health Organization

IARC SCIENTIFIC PUBLICATIONS



BIAS ASSESSMENT IN CASE–CONTROL AND COHORT STUDIES FOR HAZARD IDENTIFICATION

EDITED BY AMY BERRINGTON DE GONZÁLEZ,
DAVID B. RICHARDSON, AND
MARY K. SCHUBAUER-BERIGAN

IARC SCIENTIFIC
PUBLICATION NO. 171
STATISTICAL METHODS IN CANCER RESEARCH
VOLUME V

International Agency for Research on Cancer
World Health Organization

Suggested citation.
Berrington de González A, Richardson DB, Schubauer-Berigan MK, editors (2024). Statistical methods in cancer research, Volume V. Bias assessment in case–control and cohort studies for hazard identification (IARC Scientific Publications No. 171). Lyon, France: International Agency for Research on Cancer. Available from: https://publications.iarc.who.int/634. Licence: CC BY-NC-ND 3.0 IGO.

Sales, rights and permissions.
To purchase print copies distributed by WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, see https://www.who.int/publications/book-orders. Tel.: +41 22 791 3264; Fax: +41 22 791 4857; email: bookorders@who.int.

To obtain IARC publications in electronic format, see the IARC Publications website (https://publications.iarc.who.int).

To submit requests for adaptations or commercial use and queries on rights and licensing, see the IARC Publications website (https://publications.iarc.who.int/Rights-And-Permissions).

IARC Library Cataloguing-in-Publication Data

Names: Berrington de González, Amy, editor. | Richardson, David B., editor. | Schubauer-Berigan, Mary K., editor.
Title: Statistical methods in cancer research volume v: bias assessment in case–control and cohort studies for hazard identification / edited by Amy Berrington de González, David B. Richardson, Mary K. Schubauer-Berigan.
Description: Lyon: International Agency for Research on Cancer, 2024. | Series: IARC scientific publications, ISSN 0300-5085; v. 171. | Includes bibliographical references.
Identifiers: ISBN 9789283222293 (pbk.) | ISBN 9789283222286 (ebook)
Subjects: MESH: Neoplasms. | Bias. | Case-Control Studies. | Cohort Studies.
Classification: NLM W1

About the cover: The background image (part of the painting “Relief-disques” by Robert Delaunay, 1936) reflects the often-used depiction in the epidemiological literature of biases as scattered points (estimates) in relation to target centres (true values).

Table of contents

Contributors ix
Acknowledgements xiii
Abbreviations xv
Foreword xvii

Preface 1

Chapter 1. The role of epidemiology in cancer hazard identification by the *IARC Monographs* programme 5
1.1 Overview of cancer hazard identification in the *IARC Monographs* programme 6
1.2 Methods for evaluating human cancer studies in cancer hazard identification 9
1.3 Examples of current approaches to bias consideration in *IARC Monographs* evaluations 12
1.4 Minimizing conflicts of interest in cancer hazard identification 15
1.5 False-positives and false-negatives in cancer hazard identification: the *IARC Monographs* experience over more than 50 years 16
1.6 Conclusion 19
References 21

Chapter 2. Causal diagrams to evaluate sources of bias 23
2.1 Introduction 24
2.2 Causal DAGs to evaluate sources of bias 24
2.3 Example: building a DAG for opium consumption and lung cancer 39
2.4 DAGs and specific sources of bias 41
2.5 DAGs and multiple sources of bias 51
2.6 Signed DAGs 53
2.7 Use of DAGs in evidence synthesis 56
2.8 Summary 57
References 60

Chapter 3. Confounding: a routine concern in the interpretation of epidemiological studies 63
3.1 Introduction 64
3.2 Evaluating control for confounding 65
3.3 Tools for assessing bias due to confounding 73
3.4 Summary 82
References 84

Chapter 4. Information bias: misclassification and mismeasurement of exposure and outcome 87
4.1 Introduction 88
4.2 Qualitative evaluation of the direction of bias due to errors in exposures 90
4.3 Tools for quantifying bias due to errors in exposure 100
4.4 Outcome misclassification 110
4.5 Summary 116
References 119

Chapter 5. Selection bias and other miscellaneous biases 123
5.1 Introduction 124
5.2 Identifying selection bias in cohort studies 125
5.3 Identifying selection bias in case–control studies 132
5.4 Tools for assessing and adjusting for selection bias 139
5.5 Other miscellaneous biases 149
5.6 Summary 154
References 155

Chapter 6. Incorporating bias assessments into evidence synthesis 159
6.1 Introduction 160
6.2 Frameworks for incorporating bias assessment into evidence synthesis 161
6.3 Developing the bias-review process 162
6.4 Methods for studying multiple biases 167
6.5 Summary 172
References 173

Chapter 7. Study reporting considerations to facilitate quantitative bias assessment with access to original data 175
7.1 Introduction 176
7.2 Reporting considerations to aid graphical approaches to identify biases 177
7.3 Confounding 179
7.4 Information bias due to exposure and outcome misclassification 184
7.5 Selection bias 189
7.6 Conclusions 199
References 200

Annex 1. Evolution of the *IARC Monographs* Preamble from early investigations and reviews in the 1960s until the present day 203
A1.1 The beginnings: cancer in occupational groups 204
A1.2 Tobacco smoking and the emergence of new epidemiological methods 204
A1.3 1972: the first *IARC Monographs* 205
A1.4 1972–1980: *IARC Monographs* Volume 17 and Supplement 1 206
A1.5 1981–1990: *IARC Monographs* Supplements 4 and 7 206

A1.6 1991–2010 207
A1.7 2011 until today 207
References 208

Annex 2. Examples for which worked spreadsheets or R code are provided 211

Annex 3. Incorporating bias assessments into evidence synthesis 213
A3.1 Multiple-bias analysis: worked examples 214
A3.2 Overview of multiple-bias analysis of the data of Aliramaji et al. (2015) using pseudo-data and bias-factor approaches 217
A3.3 Sensitivity analysis 226
A3.4 A potential probabilistic multiple-bias analysis strategy 226
References 228

Index of examples 229

Index of tools 241

Disclosures of interests 243

Contributors

Editors

Amy Berrington de González
The Institute of Cancer Research
London, United Kingdom

David B. Richardson
University of California, Irvine
Irvine, CA, USA

Mary K. Schubauer-Berigan
International Agency for Research on Cancer
Lyon, France

Workshop Participants

Onyebuchi A. Arah
(unable to attend)
University of California, Los Angeles
Los Angeles, CA, USA

Laura Beane Freeman
United States National Cancer Institute
Bethesda, MA, USA

Terry Boyle
University of South Australia
Adelaide, SA, Australia

Sadie Costello
University of California, Berkeley
Berkeley, CA, USA

Veronika Deffner
University of Munich
Munich, Germany

Matthew Fox
Boston University
Boston, MA, USA

Laurence S. Freedman
(attended remotely)
Gertner Institute for Epidemiology and Health Policy Research
Tel Hashomer, Israel

Lin Fritschi
Curtin University
Bentley, WA, Australia

Irina Guseva Canu
Centre universitaire de médecine générale et santé publique
Lausanne, Switzerland

Jay S. Kaufman
McGill University
Montreal, QC, Canada

Alexander Keil
United States National Cancer Institute
Bethesda, MA, USA

Kaitlin Kelly-Reif
National Institute for Occupational Safety and Health
Cincinnati, OH, USA

Manolis Kogevinas
Barcelona Institute for Global Health (ISGlobal)
Barcelona, Spain

Hans Kromhout
Utrecht University
Utrecht, The Netherlands

Deborah A. Lawlor
University of Bristol
Bristol, United Kingdom

Sarah Lewis
University of Bristol
Bristol, United Kingdom

Ruth M. Lunn
National Institute of Environmental Health Sciences
Hillsborough, NC, USA

Brigid M. Lynch
Cancer Council Victoria
Victoria, Australia

Richard MacLehose
University of Minnesota
Minneapolis, MN, USA

Ellen Aagaard Nøhr
University of Southern Denmark
Odense, Denmark

Marie-Elise Parent
Institut national de la recherche scientifique (INRS)
Quebec City, QC, Canada

Neil Pearce
London School of Hygiene and Tropical Medicine
London, United Kingdom

Lorenzo Richiardi
University of Turin
Turin, Italy

Rodolfo Saracci
International Agency for Research on Cancer (retired)
Lyon, France

David Savitz
Brown University
Providence, RI, USA

Pamela Shaw
Kaiser Permanente Washington Health Research Institute
Seattle, WA, USA

Kyle Steenland
Emory University
Atlanta, GA, USA

Sonja Swanson
University of Pittsburgh
Pittsburgh, PA, USA

Eric J. Tchetgen Tchetgen
Wharton University of Pennsylvania
Philadelphia, PA, USA

Scott Weichenthal
McGill University
Montreal, QC, Canada

IARC Secretariat

Nathan DeBono, *IARC Monographs* Programme[1]
Pietro Ferrari, Nutrition and Metabolism Branch
Leslie Stayner, *IARC Monographs* Programme[2]
Vivian Viallon, Nutrition and Metabolism Branch
Roland Wedekind, *IARC Monographs* Programme

Additional Contributors

The following people peer-reviewed the chapters but did not participate in the Workshop or the drafting or finalization of the chapters.

John Bailer
Miami University (emeritus)
Oxford, OH, USA

Thomas Bateson
United States Environmental Protection Agency
Washington, DC, USA

Nicolas Chartres
University of California, San Francisco
San Francisco, CA, USA

Robert D. Daniels
National Institute for Occupational Safety and Health
Cincinnati, OH, USA

Sander Greenland
University of California, Los Angeles
Los Angeles, CA, USA

Neela Guha
California Environmental Protection Agency
Sacramento, CA, USA

Per Gustavsson
Karolinska Institutet
Solna, Sweden

Michael Hauptmann
Brandenburg Medical School Theodor Fontane
Neuruppin, Germany

Clare Infante-Rivard
McGill University
Montreal, QC, Canada

John Kaldor[3]
International Agency for Research on Cancer
Lyon, France

Dana Loomis
University of Nevada
Reno, NV, USA

Jay H. Lubin
United States National Cancer Institute
Bethesda, MA, USA

Elisa Pasqual
International Agency for Research on Cancer
Lyon, France

[1] Nathan DeBono is currently employed by Ontario Health and the University of Toronto, Canada.
[2] Leslie Stayner is currently employed by the University of Illinois Chicago, USA.
[3] John Kaldor is currently employed by the University of New South Wales, Australia.

Elizabeth Radke-Farabaugh
United States Environmental Protection Agency
Washington, DC, USA

Jonathan M. Samet
Colorado School of Public Health
Aurora, CO, USA

Jack Siemiatycki
University of Montreal and McGill University
Montreal, QC, USA

Administrative Assistance

Jennifer Nicholson
IARC Monographs Programme Secretary

Sandrine Ruiz
Documentalist

Production Team

Alison Lees
Scientific Editor

Karen Müller
Managing Editor

Sylvia Lesage
Principal Information Assistant

Niree Kraushaar
Project Assistant

Solène Quennehen
Senior Project Assistant

The Workshop participants and Secretariat for the Scientific Workshop on Epidemiological Bias Assessment in Cancer Hazard Identification, which met in Lyon, France, on 17–21 October 2022.

Acknowledgements

The planning of this volume in the International Agency for Research on Cancer (IARC) Statistical Methods in Cancer Research series was begun in 2019, by Dr Mary Schubauer-Berigan from the *IARC Monographs* programme in collaboration with Professor Amy Berrington de González, who was then at the United States National Cancer Institute. In 2021, Professor David Richardson was supported by IARC as a Senior Visiting Scientist to co-lead the launch of this project. Since then, many people have contributed to the development of this IARC Scientific Publication. IARC convened a group of 37 epidemiologists and statisticians who participated in a scientific workshop on epidemiological bias assessment in hazard identification, which was held in Lyon, France, on 17–21 October 2022 (see the Contributors list). During that meeting, initial drafts of the chapters were circulated and reviewed, and the group had valuable discussions for orientating subsequent developments. The book was also reviewed by several external peer reviewers, to whom the editors are exceedingly grateful. The editors greatly appreciate the skilled assistance of Ms Jennifer Nicholson in facilitating the workshop arrangements, of Ms Solène Quennehen in producing the volume, and of the volume's scientific editor, Dr Alison Lees. This project would not have been possible without the financial support of the United States National Cancer Institute and IARC.

Abbreviations

BMI	body mass index
CI	confidence interval
COI	conflict of interest
COPD	chronic obstructive pulmonary disease
CRC	colorectal cancer
DAG	directed acyclic graph
DDT	dichlorodiphenyltrichloroethane
DECOS	Dutch Expert Committee on Occupational Safety
EPIC	European Prospective Investigation into Cancer and Nutrition
ESLC	evidence suggesting lack of carcinogenicity
g-method	generalized method
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
HNSCC	head and neck squamous cell carcinoma
HR	hazard ratio
HRT	hormone replacement therapy
IARC	International Agency for Research on Cancer
ICD	International Classification of Diseases
IPAW	inverse probability of attrition weight
IRIS	Integrated Risk Information System
IV	instrumental variable
JEM	job-exposure matrix
MC-SIMEX	simulation extrapolation for misclassification
NCE	negative control exposure
NCO	negative control outcome
NHS2	Nurses' Health Study II
NRQ	non-response questionnaire
OPEN study	Observing Protein and Energy Nutrition study

OR	odds ratio
Pan Am	Pan American World Airways
PFOA	perfluorooctanoic acid
RD	risk difference
RF-EMF	radiofrequency electromagnetic field
RoC	Report on Carcinogens
RR	relative risk
SES	socioeconomic status
SIMEX	simulation extrapolation
SMR	standardized mortality ratio
SNP	single nucleotide polymorphism
SWIG	single-world intervention graph
TCDD	2,3,7,8-tetrachlorodibenzo-*p*-dioxin
WHO	World Health Organization

Foreword

Four decades ago, the International Agency for Research on Cancer (IARC) published the first volumes in the seminal Statistical Methods in Cancer Research series. Volumes I and II, edited by Dr Norman E. Breslow and Dr Nicholas E. Day, encapsulated the growing statistical methodology for case–control studies and cohort studies – the workhorse study designs (then and now) in observational epidemiology. These two volumes were instrumental in the training and education of several generations of epidemiologists, providing essential research methodologies for the study of cancer and other chronic diseases. Studies carried out using the methods outlined by Breslow and Day have formed the basis of much of our understanding about the preventable causes of cancer in humans, as evidenced in the evaluations carried out by the *IARC Monographs* programme over the decades.

Today, researchers and those involved in evidence synthesis are increasingly called upon to formally examine sources of bias in observational epidemiology studies. Although methodologies in assessing the direction and magnitude of such biases have advanced since the initial publication, there has been a dearth of accessible information available to the research and evidence synthesis communities.

In this new volume of the IARC Statistical Methods in Cancer Research series, Dr Amy Berrington de González, Dr David B. Richardson, and Dr Mary K. Schubauer-Berigan, together with their colleagues, provide a comprehensive compendium of approaches and methods with many worked examples to examine the impacts of biases in epidemiological studies. The authors are to be commended for creating a volume that lives up to the spirit and scope of its predecessors. This new volume will undoubtedly serve as a useful resource for forthcoming generations of epidemiologists and the *IARC Monographs* programme.

Dr Elisabete Weiderpass
Director, International Agency for Research on Cancer

Preface

Amy Berrington de González, David B. Richardson, and Mary K. Schubauer-Berigan

Purpose of the book

Observational epidemiology is used to identify the causes of cancer and other chronic diseases, to determine the effectiveness of interventions, and to understand reasons for differences in disease rates over time or across locations. For more than 50 years, the *IARC Monographs on the Identification of Carcinogenic Hazards to Humans* have led reviews of observational epidemiology, and other evidence, to identify preventable causes of human cancer. In this review process, expert groups in the *IARC Monographs* programme and similar programmes must judge whether a causal interpretation is supported, including whether chance, bias, and confounding can be reasonably ruled out.

Even with the best study design and analysis, it is nearly impossible to eliminate all sources of systematic bias in observational epidemiology; residual confounding, information bias, and selection bias will often remain, despite the researchers' best efforts. For cancer hazard identification, the primary concern when assessing these systematic biases is whether the direction and magnitude of the bias in the central estimate of association could change the interpretation of the result.

The primary purpose of this book is to summarize the wide range of practical methods that can be used by a reader or reviewer of a publication to assess the potential impact of confounding, information bias (including differential and non-differential exposure and outcome misclassification), or selection bias on the results of an epidemiological study. The methods we present can be implemented with information from the publications or external sources, and do not need the original study data. They include indirect approaches, for example negative control outcomes or exposures and proxies, and other approaches, such as sensitivity analyses.

The original volumes in this IARC Statistical Methods in Cancer Research series, by Breslow and Day, summarized the methods available at the time for the design and analysis of case–control studies and cohort studies. Since these works were published, in the 1980s, there have been important developments in both direct and indirect methods for identifying and quantifying biases, and related advances in causal inference. These methods are scattered across the epidemiological and statistical literature or embedded within more technical textbooks. Our goal here is to draw them together and, to quote Breslow and Day, "to place these new tools in the hands of the practising statistician or epidemiologist" (Breslow and Day, 1980, p. 7). To do this, we present them in a way that is accessible to epidemiologists and other research workers who do not have extensive statistical training, as well as to statisticians who do not have epidemiological training. We then illustrate the methods with practical examples, taken from cancer epidemiology, that recur throughout the chapters. We draw on four agents that have previously been evaluated for carcinogenicity in the *IARC Monographs* programme: radiofrequency electromagnetic field (non-ionizing) radiation, consumption of red meat, night shift work, and consumption of opium. These were chosen because they have features that illustrate the range of concepts being explored throughout the book. We provide links to online code or spreadsheets developed by the coauthors or provided by Fox et al. (2021).

Another purpose of the book is to outline the process for integrating these bias assessments into the evidence synthesis. In systematic reviews, such as those undertaken

by the *IARC Monographs* programme, biases are typically first evaluated within an individual study, and then the integration is performed. A growing range of tools is available for the appraisal of biases in systematic reviews (e.g. Grading of Recommendations, Assessment, Development, and Evaluations [GRADE], Risk of Bias in Non-randomized Studies of Interventions [ROBINS-I] or of Exposure [ROBINS-E]). Many take an algorithmic or checklist approach, which emphasizes the presence or absence of bias without regard for its direction or magnitude, and then exclude studies deemed to have the potential for (or risk of) bias (Steenland et al., 2020). A serious limitation of these tools is that they can purge many potentially informative studies, including studies that could help assess biases. On the opposite end of the evidence synthesis spectrum is the goal of reviewing and synthesizing all informative epidemiological studies. The process we outline uses the wide array of methods described in the book to retain all informative studies. This approach is consistent with the review methods described in the Preamble to the *IARC Monographs* (IARC, 2019 and Chapter 1), which calls for Working Groups to integrate studies into evidence synthesis on the basis of their quality and informativeness but recommends against the use of checklists to assess biases and sources of error.

Despite the many developments in the field of bias assessment, in many epidemiological study papers we still find the ubiquitous limitations section that acknowledges the possibility of residual confounding, measurement error, recall bias, or other deficiencies but does not attempt to assess their potential impact on findings. We hope that this book will encourage authors to apply a wider range of direct and indirect bias assessments in their primary research publications. We also refer the reader who is interested in more involved methods, including multidimensional and probabilistic bias analyses, to Fox et al. (2021). Broader adoption of these analyses will enhance the quality of the original papers and further improve the interpretation of the evidence in subsequent reviews.

This IARC Scientific Publication was supported by a 4-day workshop held in October 2022 in Lyon, France, attended by the coauthors. Before the workshop, participants developed initial literature reviews and outlined draft chapters. At the workshop, participants discussed the methods, developed worked examples, and finalized the organization of the material. The draft chapters were reviewed internally and by a group of external peer reviewers.

Definitions of biases in observational cancer epidemiology

Brief descriptions of bias in measures of association are presented next and are then further elaborated within the relevant chapters. Three major sources of systematic bias are recognized in estimates of a measure of association: confounding, information bias, and selection bias. Because the focus here is on hazard identification, rather than risk assessment, it is critical to evaluate the direction of the bias in relation to the direction of the observed effect. Therefore, we have used the terminology of bias towards or away from the null (no effect) to describe the direction wherever possible. In some special circumstances, we may deviate from this, particularly if the direction of the bias is (nearly) always upwards (positive) or downwards (negative), such as for the healthy worker effect.

Confounding

Confounding is bias that arises when the exposure and the outcome of interest share a common cause (Hernán and Robins, 2006). Confounding is a routine concern in observational epidemiology, because of the lack of random assignment to exposure that would ensure that extraneous factors (e.g. other causes of cancer) are randomly distributed among those with different exposure values. To be a confounder, a factor must be related to both exposure (the agent of interest) and outcome (the cancer of interest). Confounding can lead to spurious associations (away from the null, also termed positive confounding) or mask true associations (towards the null, also termed negative confounding). Confounding can be controlled for or minimized in the design or analysis phase of an observational study. This often requires the identification and specification of confounding factors (or confounders) that well represent the source of the potential confounding. For example, confounding by tobacco smoking may be specified in various ways, such as number of years of tobacco smoking, intensity of tobacco smoking, or time since quitting (or any combination thereof). Importantly, the quality of the specification of the confounder can influence the extent to which confounding

is controlled. Control for poorly specified confounders may result in incompletely controlled (or residual) confounding. Conversely, adjusting for a confounder that is on the causal pathway between an exposure and a cancer would have the effect of removing some of the effect of that exposure on the cancer and would give an inaccurate assessment of the true total causal effect.

Information bias

Information bias results from mismeasurement or misclassification of exposure or outcome. The extent of exposure mismeasurement or misclassification can be non-differential or differential with respect to outcome status (e.g. those with cancer can have equally accurate, more accurate, or less accurate exposure measurement or classification, compared with those without cancer). This mismeasurement can be systematic (e.g. always higher than the true value) or random (e.g. sometimes higher and sometimes lower than the true value). An example of differential and systematic exposure measurement error is the recall bias that may be observed in case–control studies, in which participants in the case group may be more likely to recall an exposure than participants in the control group, and the control group would therefore have systematically underestimated exposures. Likewise, outcome misclassification can be non-differential or differential with respect to exposure status, although the latter is less common in most observational epidemiology studies.

Selection bias

Selection bias can occur when entry into or retention in a study is related to both exposure and outcome, for example in a cohort study when exposed individuals are systematically more (or less) likely to be found to be diagnosed with cancer compared with unexposed individuals, or when dropout from a cohort is related to both exposure and outcome status. In a case–control study, selection bias can occur when people with cancer (case participants) are more (or less) likely to agree to take part in a study if they have had an exposure that they think might be related to cancer. Importantly, selection bias requires that study inclusion is related to both exposure and outcome. Study inclusion that is related to only one of these factors does not necessarily lead to selection bias. For example, if the source population giving rise to the study population is more highly exposed than the target (e.g. general) population but inclusion is unrelated to cancer outcomes, then the study might suffer from non-representativeness of the target population, but the estimate will not, in expectation, be biased for the source population. This is often the case with occupationally exposed (source) populations, who may have higher exposures than the general (target) population but whose mechanisms of follow-up would not be different from those of the general population. More information about these concepts is available in Richiardi et al. (2013).

Other bias descriptors

Other terms are used to discuss bias in epidemiological studies, although such terms often relate to problems of confounding, information bias, or selection bias.

Immortal time bias occurs when study participants (e.g. in a cohort study) cannot experience the outcome during some periods of their follow-up after exposure begins. This is usually related to the establishment of a cohort (and, hence, the start of follow-up) after the start of exposure, as might occur in occupational or pharmaco-epidemiological studies. Because immortal time bias occurs in studies that condition on disease status during some period after exposure begins, it is a form of selection bias.

Reverse causation, for example in which diagnosis with disease at time 0 causes a change in exposure status at time 1, typically refers to a type of information bias that arises when subjects are not classified with respect to baseline exposure status. Protopathic bias, which is related to reverse causation, occurs when prediagnostic symptoms of the outcome under study affect the exposure. For example, in Volume 126 of the *IARC Monographs* (IARC, 2021), protopathic bias and reverse causation were of concern for opium consumption and certain cancers, because consumption of opium (a narcotic, antitussive drug) might have been initiated to reduce early symptoms of cancers of the larynx, lung, or oesophagus.

Organization of the book

We have made some pragmatic decisions about the organization of the material in this book. The detailed index should facilitate the location of specific topics and relevant worked examples.

In Chapter 1, we provide background on the *IARC Monographs* programme and its systematic review and evidence synthesis process, as a key example of issues faced by expert review groups. We briefly discuss other major programmes of cancer hazard identification worldwide and their similarities to and differences from the processes of IARC. We also introduce the concept of study informativeness (the ability for a study to correctly identify a real positive association or a real null association) and discuss the related topic of conflicts of interest and how these could affect the potential for study bias or informativeness.

In Chapter 2, we introduce the concept of directed acyclic graphs and describe how they can be useful tools for identifying sources of bias, particularly confounding and selection bias. In Chapters 3, 4, and 5, we summarize methods that can be applied to assess and quantify the three major sources of bias (confounding, information bias, and selection bias). Chapter 5 also covers the miscellaneous topics of immortal time bias, protopathic bias, reverse causation, and considerations when using biomarkers of exposure. We then describe how to integrate these bias assessments into the evidence synthesis process in Chapter 6, and discuss some approaches for the evaluation of multiple biases. We conclude, in Chapter 7, with some recommendations for reporting results and data elements in original study publications that could facilitate bias assessment for future systematic reviewers.

References

Breslow NE, Day NE (1980). Statistical methods in cancer research, Volume I. The analysis of case–control studies (IARC Scientific Publications No. 32). Lyon, France: International Agency for Research on Cancer. Available from: https://publications.iarc.who.int/175.

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

Hernán MA, Robins JM (2006). Estimating causal effects from epidemiological data. J Epidemiol Community Health. 60(7):578–86. doi:10.1136/jech.2004.029496 PMID:16790829

IARC (2019). Preamble to the *IARC Monographs* (amended January 2019). Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

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

Richiardi L, Pizzi C, Pearce N (2013). Commentary: Representativeness is usually not necessary and often should be avoided. Int J Epidemiol. 42(4):1018–22. doi:10.1093/ije/dyt103 PMID:24062290

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

Chapter 1. The role of epidemiology in cancer hazard identification by the *IARC Monographs* programme

1.1 Overview of cancer hazard identification in the *IARC Monographs* programme 6
1.2 Methods for evaluating human cancer studies in cancer hazard identification 9
1.3 Examples of current approaches to bias consideration in *IARC Monographs* evaluations 12
1.4 Minimizing conflicts of interest in cancer hazard identification 15
1.5 False-positives and false-negatives in cancer hazard identification: the *IARC Monographs* experience over more than 50 years 16
1.6 Conclusion 19

CHAPTER 1.

The role of epidemiology in cancer hazard identification by the *IARC Monographs* programme

Mary K. Schubauer-Berigan and Rodolfo Saracci

In the *IARC Monographs* programme, epidemiological evidence is typically synthesized according to precepts that take into account whether the design, conduct, and interpretation of such studies supports a causal interpretation of their findings. Evidence syntheses can in turn be used to support various public health measures, including hazard identification, risk assessment, intervention development, and impact assessment.

Since its inception, the aim of the *IARC Monographs* programme has been to identify carcinogenic hazards for humans, by integrating, for each agent under investigation, all available evidence from studies in humans, animals, and in vitro systems. Therefore, it is important that reviewers of the evidence on cancer in humans are acquainted with the wider context of their review work. For this purpose, Section 1.1 provides an overview of the working methods and procedures used in producing the *IARC Monographs*, as applied today (Annex 1 outlines their evolution since the programme's origins), Sections 1.2 and 1.3 deal specifically with cancer epidemiology, discussing the use and evaluation of studies on human cancers with actual examples from the *IARC Monographs* programme. Section 1.4 discusses issues related to conflicts of interest (COIs). Section 1.5 examines the cancer hazard classification of all agents hitherto evaluated by the *IARC Monographs* programme from the perspective of false-positive and false-negative conclusions. Approaches for further enhancing the incorporation of bias assessments in the context of cancer hazard identification are described in the Preface and Chapter 6.

1.1 Overview of cancer hazard identification in the *IARC Monographs* programme

As the cancer research arm of the World Health Organization (WHO), IARC has estimated that there were 19.3 million new cases of cancer globally in 2020, with a projected increase of nearly 50% by 2040 (Sung et al., 2021). While part of this increase is attributable to the ageing of global populations and increasing capabilities for and access to diagnosis, particularly in low- and middle-income countries, a growing prevalence of exposure to external causes of cancer – both known and unknown – has also been postulated. Primary prevention requires identification of the causes of cancer. Since 1971, the *IARC Monographs* programme has

convened experts in cancer epidemiology, cancer bioassays, and mechanistic studies to review and evaluate the evidence on carcinogenicity of a diverse set of agents, including chemicals, particles and fibres, physical and biological agents (e.g. ionizing radiation, viruses), pharmaceuticals, complex mixtures (e.g. air pollution), personal behaviours (e.g. tobacco smoking, opium consumption), and occupational exposure circumstances (e.g. occupational exposure as a painter or as a firefighter).

Over the course of 52 years, 136 meetings have been convened of expert Working Groups to deliberate on the evidence, resulting in the publication of detailed evidence evaluations that have identified 546 agents as *carcinogenic to humans*, *probably carcinogenic to humans*, or *possibly carcinogenic to humans*. The available literature is summarized and synthesized into *IARC Monographs* volumes using an approach documented in each volume in a Preamble (IARC, 2019), which has been included since the first volume was published in 1972. As scientific methods have evolved, the Preamble has been updated accordingly (Baan and Straif, 2022), 10 times between 1977 and 2019 (see Annex 1). The Preamble lays out the steps for selecting meeting participants, for the prevention and management of COIs, and for the conduct of the meeting, as well as the methods to be used for the evidence synthesis and integration. The most recent update to the Preamble emphasizes increased transparency and scientific rigour of the review, as well as modernized methods for literature searching and screening (Samet et al., 2020), as described in Section 1.2. Side Box 1.1 provides an overview of the current evidence synthesis and integration approach.

The *IARC Monographs* programme is a process of cancer hazard identification. Working Groups ascertain whether evidence supports a causal interpretation of any observed associations between an agent and one or more types of cancer; however, they do not conduct a full risk assessment, in which the quantification of the risk of cancer associated with specific routes and levels of exposure is carried out (Samet et al., 2022). Given this focus on hazard identification, the key question faced by Working Groups is whether associations that are observed support a causal interpretation, rather than being an artefact of poor study design, the result of incorrect analysis or interpretation, or due to confounding or biases such as information bias or selection bias. The approaches by which Working Groups judge cancer epidemiology studies, individually and collectively, are described in detail in Sections 1.2–1.3. For cancer in humans, there are prespecified categories for classifying the evidence evaluation: *sufficient*, *limited*, *inadequate*, and *evidence suggesting lack of carcinogenicity* (*ESLC*). The cancer sites for which there is judged to be *sufficient* evidence have been specifically identified for each agent since *IARC Monographs* Volume 98 (IARC, 2010), while the cancer sites for which there is judged to be *limited* evidence or *ESLC* have been identified since *IARC Monographs* Volume 100 (IARC, 2012a, b, c, d, e, f). For agents suspected to cause cancer, it is not possible to design ethical experiments in humans. Consequently, most of the epidemiological evidence evaluated in the *IARC Monographs* derives from observational studies. In order to reach a determination that there is *sufficient* evidence that an agent causes cancer in humans, the Working Group judges that a causal relation has been established for one or more cancer sites, in that a positive association has been observed in the body of evidence, and that chance, bias, and confounding can be ruled out with reasonable confidence as an explanation for these positive findings. When a determination is made that the evidence is *limited*, this implies that a causal interpretation is credible, in that a positive association between exposure and cancer has been observed in the body of evidence, but chance, bias, or confounding, or some combination thereof, could not be ruled out with reasonable confidence. When it is determined that the evidence is *inadequate*, this implies that the ensemble of research does not permit a conclusion about a causal association. This usually reflects one of the following reasons: no data or sparse data were available, or a positive association was not observed in the body of evidence, or findings were positive but were judged to be entirely explained by chance, bias, or confounding.

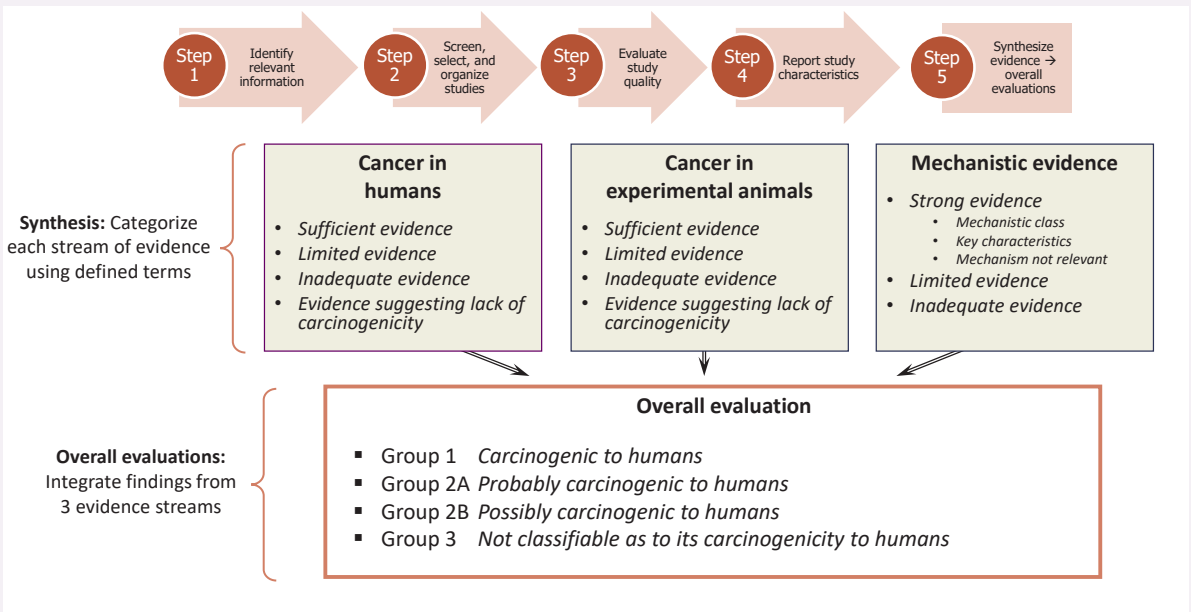
These classification categories have been largely unchanged since the revision of the *IARC Monographs* Preamble in Volume 30 in 1982 (IARC, 1983), when the phrase “chance, bias, and confounding” was introduced to differentiate between *sufficient* and *limited* evidence (see Annex 1). For a determination of *ESLC*, a judgement is made that no positive findings

Side Box 1.1. Evidence synthesis and integration in the *IARC Monographs*

As laid out in the current Preamble, adopted in 2019, the *IARC Monographs* evaluations are carried out in a five-step systematic review process (Fig. 1.1). Step 1 is the identification of relevant studies, by conducting extensive literature searches. Step 2 involves screening, selecting, and organizing the identified studies. Study quality (including consideration of potential biases) is evaluated in Step 3, and study characteristics are reported in Step 4. Step 5 of the review process, evidence synthesis and integration, is conducted at an 8-day meeting held at IARC in Lyon, France.

Three streams of evidence are considered in the *IARC Monographs* evaluation process: cancer in humans, cancer in experimental animals, and mechanistic evidence. The evidence is first synthesized individually by stream using well-defined criteria. Then the evidence is integrated across the streams, using guidelines established in the Preamble, into one of four groups: Group 1, *carcinogenic to humans*; Group 2A, *probably carcinogenic to humans*; Group 2B, *possibly carcinogenic to humans*, and Group 3, *not classifiable as to its carcinogenicity to humans* (Fig. 1.1). (text continues on page 7)

Fig. 1.1. Overview of the *IARC Monographs* evidence synthesis and evaluation process. Source: Compiled from Samet et al. (2020).



were seen in adequately powered and well-conducted studies at any exposure level and that bias could be ruled out as an explanation for the absence of an association. For example, for coffee drinking, there was deemed to be *ESLC* for cancers of the pancreas, liver, female breast, uterine endometrium, and prostate (IARC, 2018a). In practice, a designation of *ESLC* is often used when an inverse association is observed for a cancer site (e.g. such an inverse association was noted for coffee drinking and cancers of the liver and endometrium). Typically, *ESLC* for one or more cancer sites may occur together with *sufficient* or *limited* evidence of carcinogenicity for other sites (e.g. the agent tamoxifen exhibited *sufficient* evidence for causation of endometrial cancer and *ESLC* for breast cancer; IARC, 2012a).

The evidence for the two other streams, cancer in experimental animals and carcinogen mechanisms, is synthesized using different approaches from that used for cancer in humans (Samet et al., 2020). Once an evaluation is made regarding the evidence synthesis for each individual evidence stream, the three streams

Fig. 1.2. Possible combinations leading to overall evaluations during evidence integration in the *IARC Monographs* programme of cancer hazard identification.

Evidence of cancer in humans	Evidence of cancer in experimental animals	Mechanistic evidence	Evaluation
Sufficient	Irrelevant	Irrelevant	*Carcinogenic* (Group 1)
Limited or *inadequate*	*Sufficient*	*Strong* (exposed humans)	
Limited	*Sufficient*	*Limited* or *inadequate*	*Probably carcinogenic* (Group 2A)
Limited	*Limited* or *inadequate*	*Strong*	
Inadequate	*Sufficient*	*Strong* (human cells or tissues)	
Limited or *inadequate*	Irrelevant	*Strong* (mechanistic class)	
Limited	*Limited* or *inadequate*	*Limited* or *inadequate*	*Possibly carcinogenic* (Group 2B)
Inadequate	*Sufficient*	*Limited* or *inadequate*	
Inadequate	*Limited* or *inadequate*	*Strong*	
Inadequate	*Sufficient*	*Strong* (does not operate in humans)	*Not classifiable* (Group 3)
All other situations			

are integrated by the Working Group into an overall synthesis leading to one of the four classification groups (Fig. 1.2).

The overall evaluation reflects the degree of certainty about the strength of evidence regarding the carcinogenicity of the agent to humans. A determination of *sufficient* evidence regarding one or more cancer sites in humans leads directly to a Group 1 classification, regardless of the evidence in the other two streams. If *sufficient* evidence for cancer in humans is not shown for any cancer site but there is *limited* evidence regarding one or more cancer sites, then evaluations from the other two streams may inform the overall classification: a determination of either *sufficient* evidence for cancer in experimental animals or *strong* mechanistic evidence (or both) combines with the *limited* evidence for cancer in humans to give a Group 2A classification. Recent examples include night shift work and 1,1,1-trichloroethane: for night shift work, human cancer evidence was *limited* for cancers of the breast, prostate, and colorectum (IARC, 2020); for 1,1,1-trichloroethane, evidence was *limited* for multiple myeloma (IARC, 2022). In both instances, there was *sufficient* evidence for cancer in experimental animals, and for night shift work there was also *strong* mechanistic evidence in experimental systems. However, in most instances, particularly for environmental or occupational exposures, a determination of *sufficient* evidence for cancer in humans is accompanied by *sufficient* evidence in experimental animals, *strong* mechanistic evidence, or both (Cogliano et al., 2011; IARC, 2012b, c, d, e, f). It is possible to arrive at a Group 1 classification with *limited* or even *inadequate* evidence regarding cancer in humans if there is *sufficient* evidence for cancer in experimental animals and *strong* mechanistic evidence in exposed humans. Three examples are ethylene oxide (IARC, 1994), neutron radiation (IARC, 2000), and perfluorooctanoic acid (PFOA; Zahm et al., 2024).

1.2 Methods for evaluating human cancer studies in cancer hazard identification

1.2.1 The IARC Monographs *approach*

Section B.2 of the *IARC Monographs* Preamble (IARC, 2019) presents two parts specifically devoted to human cancer studies: the first details considerations in the evaluation of individual studies, and the second addresses considerations for evaluation of the overall body of evidence. The first part (Sections B.2a–B.2c) addresses the types of study to be considered for the evaluation of human cancer

evidence, indicating that high-quality case–control and cohort studies usually provide the most suitable data for such an exercise; it then mentions the procedures to be followed for the identification of eligible studies of cancer in humans and outlines the key aspects of assessment of an individual study's quality and informativeness (the latter term designating the overall ability of a study to identify an effect when one exists, or to identify the lack of an effect when none exists). Four cardinal aspects of each study should be examined: the study description and design, the study population (including subpopulations, such as people potentially susceptible to cancer), the outcome measurement, and the exposure measurement. Furthermore, in evaluating the adequacy of statistical methods of analysis, which have evolved considerably in scope and sophistication in recent decades, the role of random and systematic errors, collectively designated as chance, bias, or confounding, should be considered. The *IARC Monographs* Preamble (IARC, 2019) notes, "For the sake of economy and simplicity, in this Preamble the list of possible sources of error is referred to with the phrase 'chance, bias, and confounding', but it should be recognized that this phrase encompasses a comprehensive set of concerns pertaining to study quality." The Preamble emphasizes, "These sources of error do not constitute and should not be used as a formal checklist of indicators of study quality. The judgement of experienced experts is critical in determining how much weight to assign to different issues in considering how all of these potential sources of error should be integrated and how to rate the potential for error related to each of these considerations." As a transition to the second part, the combination of studies via meta-analysis and pooled analyses is sketched (in Section B.2d) as a valuable, albeit not prescriptive, tool to check the consistency of results across studies.

The second part of Section B.2e presents a range of considerations in assessing the body of epidemiological evidence, stating in the opening paragraph, "There is no formulaic answer to the question of how many studies of cancer in humans are needed from which to draw inferences about causality, although more than a single study in a single population will almost always be needed. The number will depend on the considerations relating to evidence described." This part carries an obvious footprint of the viewpoints presented by Hill (1965), from which the available epidemiological evidence needs to be critically scrutinized. Although formulated to assist causal inference on environmental exposures of various kinds, the Hill perspective has become more generally influential in discussions on causal inference from observational studies. Set aside from these viewpoints is the issue of ruling out chance, namely the effects of sampling errors, estimated by tests of significance and confidence limits, on which Hill takes a firm position (Hill, 1965): "No formal tests of significance can answer [causal] questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the 'proof' of our hypothesis." This position remains, in essence, valid today (Savitz et al., 2024).

Each of these viewpoints focuses on a feature of the epidemiological data that, if present, supports a causal interpretation of an observed association between an exposure and a risk of cancer. The nine features, as labelled by Hill (but in a different order), are reported in Table 1.1. With the exception of temporality, which is, in fact, an absolute requirement, the relative weight of each feature is not fixed, and the absence of a feature does not automatically detract from a causal interpretation. However, consistency – which reflects, within an observational context, the important concept of reproducibility in science – ranks generally high in weight but must be balanced by consideration of the relative informativeness and potential for bias of the different studies contributing evidence to the evaluation.

The evaluation of several of the features in Table 1.1 to infer the causal nature of an observed association evolved early in the *IARC Monographs* programme; it has been maintained for the past 40 years, unchanged at its core, and has been accompanied by several specifications and explicit indications, outlined in Annex 1. For example, consideration number 5, consistency (which has always had a prominent role for causal inference), has recently been better specified in terms of triangulation methods (e.g. Lawlor et al., 2016; for additional information, see Chapter 6). As further examples, the Preamble now advises Working Groups to explicitly consider the direction and magnitude of biases (e.g. as arising from exposure misclassification or unmeasured confounding), not simply their presence, and discusses the possibility of *ESLC* of an exposure.

Table 1.1. Features of an association between exposure and cancer risk that support a causal interpretation within the *IARC Monographs* programme

Feature	Evaluation within *IARC Monographs*
1. Temporality	There should be unequivocal evidence that the onset of exposure has preceded the onset of a detectable cancer.
2. Strength	Once all feasible adjustments for confounding and biases have been implemented, a strong resulting association (e.g. with high relative risk) is less likely than a weak one to be fully explained by residual or unknown confounding and biases, and therefore is more likely to be of causal nature.
3. Specificity	This consideration, suggesting that evidence is stronger when carcinogenicity is observed in only one or a few organs or tissues (rather than in many), has been variably invoked by Working Groups. For agents that exhibit systemic exposure (e.g. ionizing radiation), specificity is not highly valued. For other agents, where exposure is not systemic (e.g. some lung carcinogens), a finding of specific effects only in organs where exposure occurs strengthens a causal interpretation. Furthermore, an association may sometimes be judged as much stronger when exposure is redefined by restriction to specific subgroups (e.g. people with a particular genetic polymorphism or exposed to a single chemical) or the outcome is restricted to specific histological or molecular subtypes of a cancer.
4. Biological gradient	In carcinogenesis, an all-or-none response to a carcinogen very rarely, if ever, occurs. Hence, finding that an increasing exposure level is associated with an increasing cancer risk is in accordance with established biological knowledge on cancer causation.
5. Consistency	A causal interpretation of an association receives considerable support when findings are consistent between studies carried out in different populations, with possibly different exposure and confounding patterns or effect modification, or with different study designs and methods (accounting for differences in study informativeness, e.g. exposure contrast or latency considerations). Study informativeness is an important consideration here. A study is informative to the extent that it is capable of detecting an increased risk when it truly exists; this goes beyond study power and depends on the availability of the right population with the right exposures and the right design with the right cancer type. Fully informative studies permit sounder interpretation of results than do minimally informative studies.
6. Experiment	A reduction of risk after reduction or cessation of an exposure points to the exposure as the causative agent of the risk; this indication carries particular weight if the reduction or cessation occurs in the framework of a purposely designed intervention (e.g. a regulatory measure to reduce the level of an air pollutant).
7. Plausibility	Firmly established biological mechanisms (e.g. a precursor lesion well documented as entailing a high risk of subsequent cancer) speak in favour of the causal nature of the association between an exposure and a cancer if, for example, the same exposure is also found to be associated with the precursor lesion. Biological mechanisms still under investigation do not contribute to the evaluation of the evidence in humans and are examined separately, within the mechanistic evidence stream of an *IARC Monographs* evaluation process.
8. Coherence	In Hill's words, the causal interpretation "should not seriously conflict with the generally known facts" (Hill, 1965) about the disease and – it can be added – the exposures, such as their respective distributions, patterns, and trends within and between populations. Coherent findings across related cancer sites with respect to exposure to the target organ (e.g. as for cigarette smoking or alcohol consumption) can support a causal interpretation.
9. Analogy	This weak feature is not usually considered, except in the strict sense of regarding as analogous certain chemicals with very close structural and activity properties; this consideration would occur in the mechanistic evidence stream evaluation for the *IARC Monographs*.

Side Box 1.2. Examples of other programmes of cancer hazard identification

Since 1978, the United States Department of Health and Human Services has had the legislative mandate to publish a cancer hazard report (prepared by the National Toxicology Program), known as the Report on Carcinogens (RoC), which lists substances (defined as "agents, substances, mixtures, exposure scenarios") that are "either known or reasonably anticipated to be a human carcinogen" (Lunn et al., 2022). Evaluation by the RoC requires that a significant number of people be exposed in the USA. Like the *IARC Monographs* programme, the National Toxicology Program RoC adheres to a well-defined and structured process for evaluating substances for their carcinogenic hazard. This process also includes consideration of human cancer, animal bioassay, and mechanistic evidence streams. Considerations in the evaluation of human cancer studies are similar to those used in the *IARC Monographs* (Lunn et al., 2022). Study informativeness (identified from assessments of risk of bias and study sensitivity) is emphasized for human cancer evaluations. One major difference between the programmes is that the RoC is drafted by scientific staff within the National Institute of Environmental Health Sciences programme, rather than external expert Working Groups, and goes through external expert peer review and public comment before finalization. A detailed comparison of cancer classification methods and results has been published by Lunn et al. (2022), but it is worth noting that there is generally high concordance between the agents classified in Group 1 in the *IARC Monographs* and those classified by the RoC as "known to be carcinogenic to humans".

Other major programmes that undertake hazard identification do so within the context of a formalized risk assessment, for example the Integrated Risk Information System (IRIS) programme of the United States Environmental Protection Agency (U.S. EPA, 2022), the Dutch Expert Committee on Occupational Safety (DECOS; Health Council of the Netherlands, 2012), the European Food Safety Agency, and many others. These programmes evaluate human cancer evidence in a variety of ways, often using evaluation approaches (e.g. IRIS and DECOS) similar to those of the *IARC Monographs* programme, with careful consideration of study quality and potential for bias. (text continues below)

1.2.2 Other major programmes of cancer hazard identification

While the *IARC Monographs* programme is the world's oldest cancer hazard identification programme, other health organizations worldwide have been engaged in the conduct of cancer hazard identification, some for decades. Side Box 1.2 briefly mentions a few such programmes, emphasizing the extent to which their evaluation approaches differ from those of the *IARC Monographs*.

1.3 Examples of current approaches to bias consideration in *IARC Monographs* evaluations

Historically, and specifically since the implementation of the Preamble revision in 1987 (IARC, 1987), Working Groups have used a variety of approaches to determine whether chance, bias, and confounding can be ruled out with reasonable confidence, as a delimiter between evaluations of *sufficient* and *limited* evidence, or whether a causal interpretation is even credible, in distinguishing between *limited* and *inadequate* evidence. Working Groups closely scrutinize the adequacy of study design and analysis methods and of reporting of results, noting detailed strengths and limitations of the studies evaluated. The evidence triangulation principle has long been applied in considering whether different studies that have diverse types of bias point to the same conclusion. For example, ecological and case–control studies of arsenic in drinking-water had different bias potentials from each other and from cohort studies of inhalation exposure to arsenic in workers; however, all three types of study strongly pointed to an excess risk of lung cancer (IARC, 2012c). Case–control studies of low-level radon exposure in the general population (which had some potential for recall bias and non-differential exposure misclassification) complemented cohort studies of uranium miners exposed to high-dose radiation levels, lending confidence to a causal interpretation of the association between radon progeny and lung cancer (IARC, 2012d).

Negative control outcomes, i.e. outcomes that are plausibly related to confounders but not to the agent of interest, can help elucidate whether confounding exists. As an example,

the association between an agent and chronic obstructive pulmonary disease (COPD) is often examined by Working Groups in conjunction with that observed for the agent in question and lung cancer. Because COPD is related strongly to tobacco smoking but less strongly, or not at all, to many other lung carcinogens, the absence of an association between an agent and COPD provides reassurance that smoking is not a confounder of the association observed between the agent and lung cancer. More quantitative approaches when information about the confounder is available for only some subjects (or is not available for any subject), such as the use of indirect adjustments, and worst-case assumptions about confounder–exposure distributions, have been rarely used by Working Groups but are explicitly mentioned in the Preamble (IARC, 2019, p. 17).

The current Preamble (IARC, 2019, pp. 15–16) emphasizes the explicit evaluation of exposure assessment quality, including the expected impact of any related biases on the direction and magnitude of measures of association between exposure and cancer.

To illustrate the approaches used by Working Groups, we draw examples from the four topics of interest that will be discussed throughout the rest of this volume (as noted in the Preface): radiofrequency electromagnetic field (RF-EMF) radiation (Example 1.1), consumption of red meat (Example 1.2), night shift work (Example 1.3), and consumption of opium (Example 1.4). It is important to note that, in addition to concerns about bias and confounding, study

Example 1.1. Evaluation of radiofrequency electromagnetic field (RF-EMF) radiation by the *IARC Monographs*

Radiofrequency electromagnetic fields, as generated in mobile phone use, were evaluated in *IARC Monographs* Volume 102 as *possibly carcinogenic to humans* (Group 2B), on the basis of *limited* evidence for cancer in humans and *limited* evidence of carcinogenicity in experimental animals (IARC, 2013). The Working Group noted in their rationale that the human epidemiological evidence was mixed. Some small case–control studies, several studies of occupational exposure, and a large cohort study, all investigating brain tumours (particularly gliomas) were regarded as uninformative because of several potential sources of exposure misclassification and insufficient control for possible confounding. The bulk of the evidence came from reports of the Interphone study – a very large international, multicentre case–control study – and a separate large case–control study in Sweden on acoustic neuroma and glioma and meningioma of the brain. Both studies showed an association between mobile phone use and glioma and acoustic neuroma. However, each study presented non-negligible limitations. In the Interphone study, an increased risk of glioma was found only for the highest levels of estimated cumulative exposure (cumulative call time). However, differential participation rates between participants in the case and control groups – compounded with lower participation rates of control participants who were non-regular mobile phone users than of control participants who were regular users – could have resulted in a lower estimated risk of brain cancer among regular mobile phone users than the true risk for the participating centres. This is one of the reasons given that chance and bias could not be excluded as possible explanations for the increased risk at the highest levels of exposure. The study in Sweden revealed an increased risk of glioma, with a gradient with increasing cumulative call time. The sequential approach, using a self-administered questionnaire followed by a phone interview to collect exposure and confounder information, raised the possibility of information bias, with validation studies not having been carried out in the pertinent population. There were also concerns about recall bias, which were somewhat tempered by the specificity of the positive associations for two tumour subtypes (glioma and acoustic neuroma) but not others. The limitations of the two studies led the Working Group to the evaluation that there was *limited* evidence for cancer in humans; it appears that the reviewers had made full use of the published results in the main and ancillary publications of all studies, and especially of the Interphone study and the study in Sweden, to probe the existence and direction of biases without, however, formally estimating the overall impact of biases for each study. (text continues above)

Example 1.2. Evaluation of red meat consumption by the *IARC Monographs*

Red meat consumption was evaluated as *probably carcinogenic to humans* (Group 2A), on the basis of *limited* evidence for cancer in humans and *strong* mechanistic evidence (IARC, 2018b). The Working Group identified a large number of cohort and case–control studies, conducted across five continents. They noted substantial variation in the quality of study design and exposure assessment instruments, as well as in the definition of red meat consumption. Cohort studies with quantitative information on red meat consumption derived from validated dietary questionnaires and with good control for confounding were deemed most informative, together with a small subset of case–control studies, in examining risk of colorectal cancer. The main determinant in reaching a conclusion of *limited* evidence for cancers of the colorectum and pancreas in humans was the inconsistency of results in some of the larger, higher-quality studies. For prostate cancer, concerns about reporting bias and outcome misclassification for aggressive forms of disease were additionally mentioned. No formal appraisal of bias was carried out for these or other cancer sites in relation to red meat consumption. (text continues on page 13)

Example 1.3. Evaluation of night shift work by the *IARC Monographs*

Night shift work (IARC, 2020) was evaluated as *probably carcinogenic to humans* (Group 2A), on the basis of *limited* evidence for cancer in humans, *sufficient* evidence for cancer in experimental animals, and *strong* mechanistic evidence in experimental systems. There were two types of human cancer study, with different bias concerns: cohort studies of night shift workers in the general population as well as among nurses and flight crew, and population-based case–control studies. Most cohort studies did not show positive findings, but most could not detect associations for specific time windows of sensitivity for induction of breast cancer (e.g. premenopausal breast cancer after recent or non-recent night shift work). Others had short follow-up periods, leading to concerns about study power or the ability to detect cancer risk with long latency. In addition, non-differential exposure misclassification was a serious concern, but the Working Group did not attempt to quantify the magnitude of this bias. A large and informative pooled case–control study (Cordina-Duverger et al., 2018) showed positive associations between night shift work and breast cancer overall, with a positive exposure–response association observed for only one of several exposure metrics. Here, differential exposure misclassification (due to recall bias) and selection bias were of primary concern, with bias away from the null (i.e. a no-association measure) being thought most likely, but the Working Group did not estimate the magnitude of the bias or whether it could explain the magnitude of risk elevation found in the case–control studies. The Working Group concluded that there was *limited* evidence for breast cancer (as well as cancers of the prostate and colorectum) in humans. (text continues on page 13)

Example 1.4. Evaluation of opium consumption by the *IARC Monographs*

In 2020, opium consumption was evaluated as *carcinogenic to humans* (Group 1), with *sufficient* evidence for cancer in humans (IARC, 2021). The evidence regarding an association between opium consumption and cancer consisted of one large well-conducted cohort study and several dozen case–control studies. All were population-based or hospital-based, and most were conducted in the Islamic Republic of Iran. In the Working Group's evaluation, the use of causal diagrams helped to elucidate which covariates might be confounders. The Working Group used directed acyclic graphs (DAGs) to identify the main concerns regarding bias; these included residual confounding (primarily by tobacco smoking), selection bias, and recall bias for case–control studies, and non-differential exposure misclassification, reverse causation, and protopathic bias for all study designs.

The cohort study was subject to non-differential exposure misclassification, and exposure history was captured at one time point and was not further updated. However, the use of biomarkers of opium metabolites was thought to provide good validation for the questionnaire-based exposure assessment method. Residual confounding by tobacco smoking was a second concern, although the cohort study had detailed estimates of several smoking measures, which were used to adjust for tobacco smoking. Opium-related risk was also examined in never-smokers of tobacco.

In the population-based case–control studies, the main concern was recall bias, and there was some evidence that the choice of control group influenced the estimated odds ratios. Selection bias due to differential participation rates of case and control participants was a potential concern, as were protopathic bias and reverse causation. However, the latter two sources of bias were thought to have been adequately dealt with by investigators during the analyses. A formal assessment of the impact of some of these sources of bias was evaluated in an annex (IARC, 2021), and this work was important to the Working Group's evidence synthesis, which concluded that there was *sufficient* evidence in humans that opium consumption causes cancers of the lung, larynx, and bladder, and *limited* evidence that opium consumption causes cancers of the oesophagus, pancreas, pharynx, and stomach. However, the different sources of potential bias were evaluated individually and were not combined in any quantitative analysis.
(text continues on page 13)

informativeness is used to evaluate reasons for consistency (or not) of findings, one of the key principles of causal inference used in the *IARC Monographs*.

1.4 Minimizing conflicts of interest in cancer hazard identification

In contemporary research, COI is a widespread phenomenon, but its structural social aspects and causes are beyond the scope of this volume. The relevance of COIs here stems from the potential for inducing erroneous scientific judgements in cancer hazard identification, hampering and delaying the attainment of scientifically valid evidence, with the consequence of increased health and economic costs to society. The United States Institute of Medicine defines a COI as "a set of circumstances that creates a risk that professional judgement or actions regarding a primary interest will be unduly influenced by a secondary interest" (Lo and Field, 2009). Of these circumstances, research funding, including employment support, and personal financial interests have been well documented after surveys of published studies as having the potential to distort scientific judgements in several areas of epidemiology, including studies for cancer hazard identification (Michaels, 2008; Mandrioli et al., 2016; Lundh et al., 2017). Reviewers of the evidence pertinent to cancer hazard identification are at times confronted with the situation where the influence of an identified COI on the aims, overall informativeness, design, results, and interpretation of a study cannot be directly evaluated. In such instances, separate consideration and comparison of results can be made of studies involving clear COIs and studies not so affected, with full reports on whether and why this examination leads to equal or different treatment of the results of the two types of study when drawing interpretative conclusions.

Of course, reviewers themselves may have COIs, and in evidence evaluation and synthesis COI avoidance is no less important than methodological correctness. In the *IARC Monographs* programme, IARC has developed and applies a COI prevention and control policy. Before a Working Group meeting, each potential participant, including the IARC Secretariat, fills in a WHO declaration of interests form to report financial interests, employment and consulting work (including remuneration for serving as an expert witness), individual and institutional research support, and non-financial interests, such as public statements and positions related to the meeting. The declared interests are then assessed to determine whether there is a COI barring participation in the Working Group in question. Meeting participants occupy one of five positions: Working Group full member, Invited Specialist, Representative (of a national or international health agency), Observer, or IARC Secretariat member. Only Working Group full members, assessed as having no COI, can take part in all phases of the evidence evaluation, while other participants have different limitations (Table 2 of IARC, 2019) to formally control for potential COI effects arising from their positions.

It is important for a reader of an *IARC Monograph*, papers cited in it, and published commentaries on it, to consider the possible presence of COIs by carefully examining COI and funding statements and, when in doubt, even an author's body of work beyond the single paper being consulted. The mere presence of COIs may indeed be difficult, if not impossible, to detect if no information at all is provided or when authors declare no COI despite, for instance, funds for the work being provided by the producer or user of the agent under evaluation.

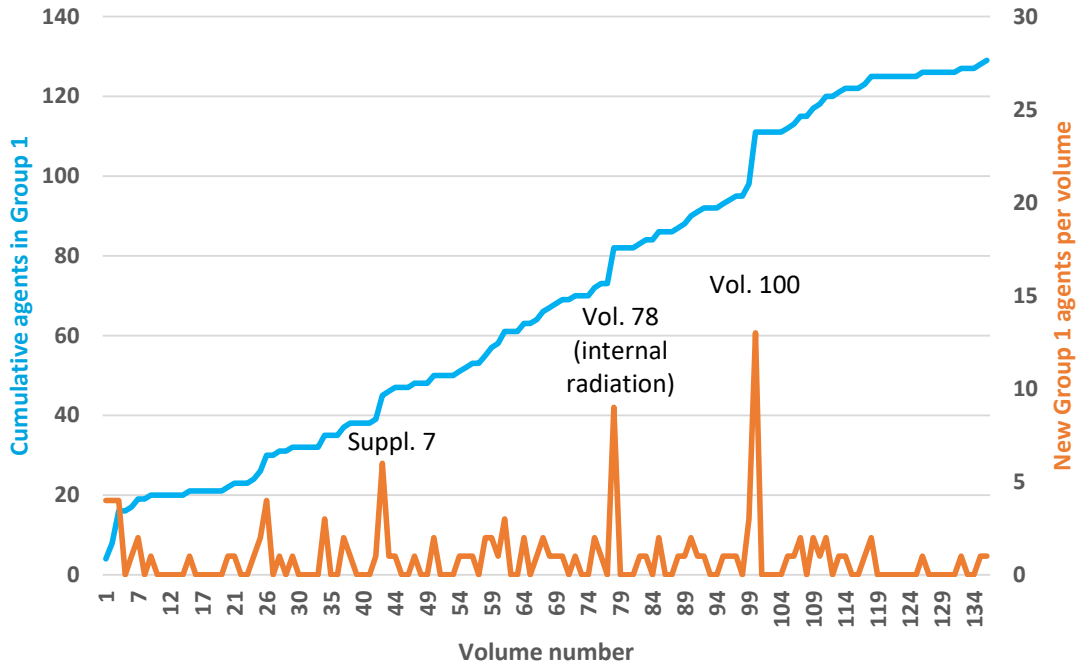
Different competent and COI-free researchers may legitimately take varying viewpoints on the same body of evidence. The Preamble instructs the IARC Secretariat to include a representation of diverse credible viewpoints when assembling a Working Group. Such diversity of viewpoints can be essential in ensuring that all aspects of study quality, informativeness, and potential for bias are brought forward for deliberation and evaluation by the Working Group; this also minimizes any risk of bias that may derive from the viewpoints of Working Group participants themselves as authors of studies of the agent being evaluated.

1.5 False-positives and false-negatives in cancer hazard identification: the *IARC Monographs* experience over more than 50 years

Agents are prioritized for evaluation in the *IARC Monographs* programme if there is evidence for human exposure and some evidence for or suspicion of carcinogenicity (IARC, 1998), based on studies in humans or animals (Samet et al., 2020). Thus far, in the *IARC Monographs* programme, 129 agents have been identified as *carcinogenic to humans* (Group 1). There has been a steady growth in the identification of these carcinogenic agents over the life of the programme (Fig. 1.3), with step changes at particular points when certain agents with an abundance of human cancer evidence were considered eligible for evaluation (e.g. biological agents in the mid-1990s, ionizing radiation in the late 1990s), and in the re-evaluations of all agents, published in *IARC Monographs* Supplement 7 (IARC, 1987), and of all Group 1 agents, published in *IARC Monographs* Volume 100 (Cogliano et al., 2011). (In *IARC Monographs* Volume 100, some broad agent groupings were divided to better denote the different cancer sites in humans with *sufficient* or *limited* evidence.)

In evaluating the human cancer evidence, as noted in Section 1.2, expert Working Groups judge whether the evidence at hand supports a causal interpretation with reasonable confidence. The question may arise about to what extent this process of judgement is likely to result in false-positives (e.g. a declaration that there is *sufficient* evidence for a causal association between the agent and a given cancer site when the association is actually not causal) or false-negatives (e.g. a failure to identify a truly causal association). Several critics have argued that expert judgement of human cancer observational data has the potential to produce many false-positives (Taubes, 1995; Ioannidis, 2005; Boffetta et al., 2008). Other authors have suggested that such concerns lack foundation, in part based on the experience of the *IARC Monographs* programme (Cogliano et al., 2004; Blair et al., 2009; Pearce et al., 2015; Saracci, 2017; McCullough et al., 2022). Over the 52-year history of the programme, there have been many opportunities to examine this question in detail. During this time, a determination that there is *sufficient* evidence of carcinogenicity in humans for at least one cancer type has almost

Fig. 1.3. Time series showing the addition of new agents in Group 1 over the 52-year history of the *IARC Monographs* programme.

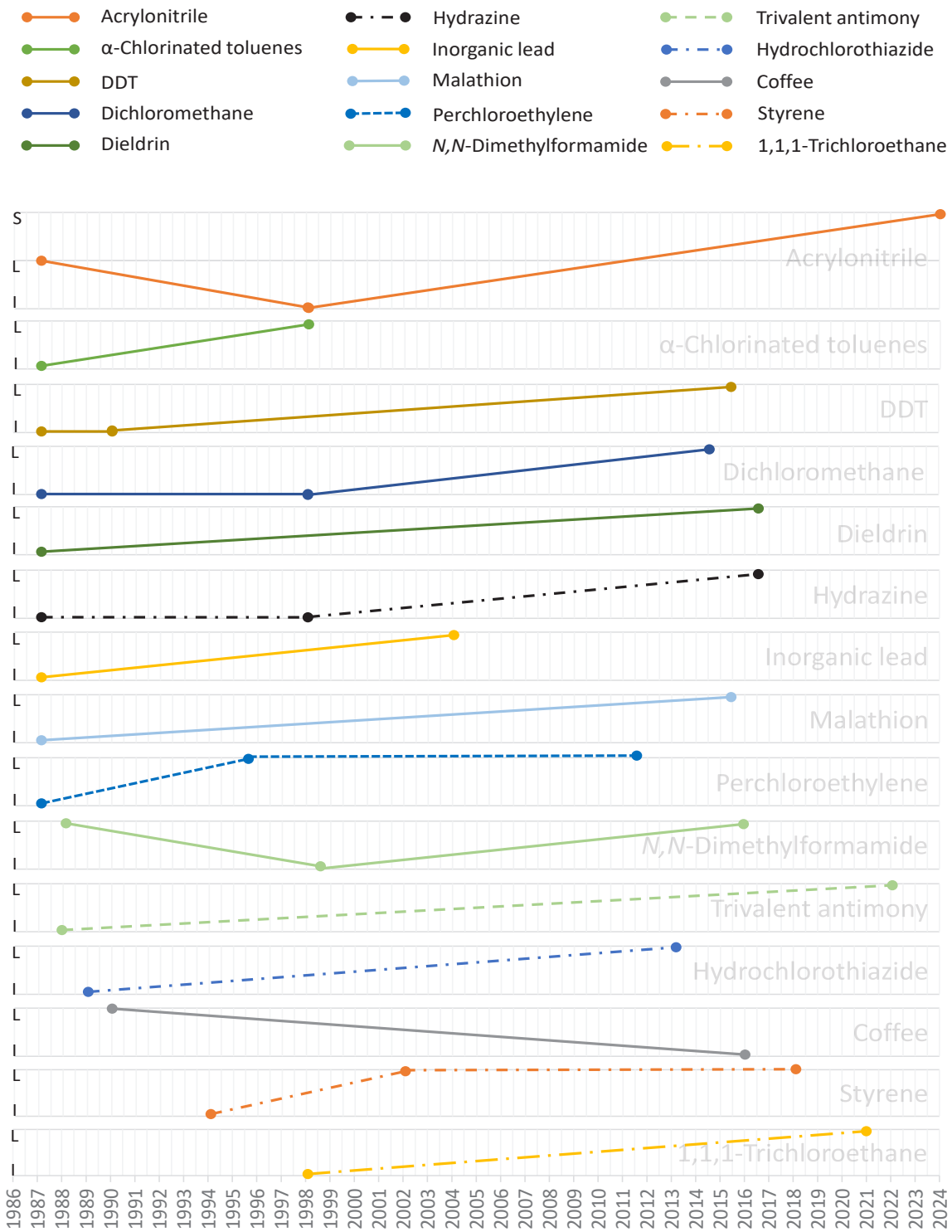


never been reversed. For example, as published in *IARC Monographs* Volume 100 (IARC, 2012a, b, c, d, e, f), different Working Groups re-evaluated the evidence for all the (more than 100) agents then classified in Group 1. With the exception of human papillomavirus (HPV) type 66, all of these re-evaluated agents were reaffirmed as Group 1. For many, if not most, of the agents, the human cancer evidence had strengthened since the previous evaluation, and additional cancer sites with *sufficient* or *limited* evidence were identified. There is also broad concordance between classifications of *sufficient* evidence in the *IARC Monographs* and those in other hazard identification programmes (Lunn et al., 2022). These findings suggest that there is a low false-positive rate for a determination that there is *sufficient* evidence of carcinogenicity in humans.

The category of *limited* evidence is characterized by some uncertainty, in which new evidence from informative studies might be expected to shift the evaluation to either *sufficient* or *inadequate* (or even *ESLC*). However, in practice, agents have more often moved upwards in classification than downwards. For agents with *limited* evidence in humans, in many cases (e.g. arsenic, dioxin, polychlorinated biphenyls, trichloroethylene) the evaluations have advanced, over time, to *sufficient*. Other agents that have moved between classifications over time were much more likely to move up from *inadequate* to *limited* than down from *limited* to *inadequate* (Fig. 1.4). Examples of such agents that have moved up include industrial chemicals, such as α-chlorinated toluenes, dichloromethane, styrene, and 1,1,1-trichloroethane, and pesticides, such as dichlorodiphenyltrichloroethane (DDT), dieldrin, and malathion. Coffee is an example of an agent that has moved down from *limited* to *inadequate*. *N,N*-dimethylformamide moved from *limited* to *inadequate* in 1998 and back to *limited* in 2016. Acrylonitrile moved from *limited* to *inadequate* in 1998 and then to *sufficient* in 2024.

In 1983, a workshop held in Oxford, United Kingdom, discussed interpretations of so-called negative evidence in human studies (i.e. evidence deriving from studies in humans that was deemed to be unconvincing) for 10 agents with *sufficient* evidence from cancer bioassays (Wald and Doll, 1985). For most of the 10 agents, the workshop attendees concluded that the evidence was

Fig. 1.4. Agents whose classification has shifted between categories of *inadequate* and *limited* regarding human cancer over the life of the *IARC Monographs* programme (excluding most agents that were eventually classified as *sufficient*). I, *inadequate*; L, *limited*; S, *sufficient*.



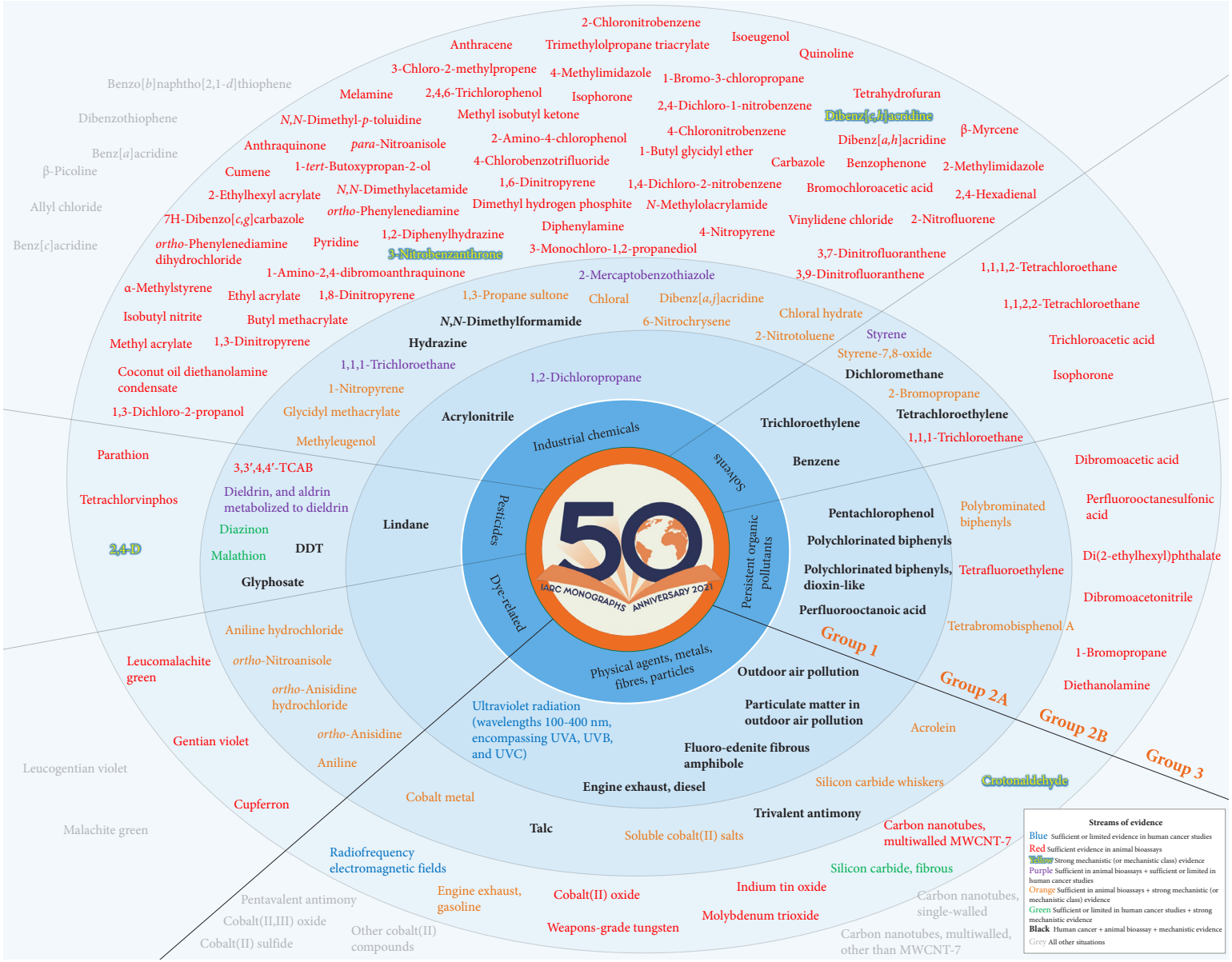
likely to remain classified as *inadequate* or even as *ESLC* in humans. Notably, in the 40 years since this workshop, 3 of the 10 agents (beryllium, formaldehyde, and oral contraceptives) were found to have *sufficient* evidence for cancer in humans, and another 4 (DDT, hydrazine, nitrites, and hairdresser exposures to dyes) to have *limited* evidence in humans. Improvements in the number, quality, and informativeness of epidemiological studies were key to these changes for these agents, whose previous evaluations could be viewed as false-negatives. A similar analysis of the agents for which there had been *inadequate* evidence regarding cancer in humans, published in *IARC Monographs* Supplement 7 (IARC, 1987), found that many of these had advanced in classification since then (Cogliano et al., 2004). Such patterns suggest that many epidemiological biases in the literature on carcinogenicity (e.g. exposure misclassification, selection biases, and even confounding) are operating in a downward direction or towards the null.

One potential reason for the relatively low false-positive rate in the classification of agents in Group 1 is the fact that several lines of evidence contribute to the nomination of agents for evaluation; in other words, potential carcinogenicity in humans is often preceded by evidence of cancer in experimental animals or of carcinogen mechanisms. For nearly all the Group 1 agents re-evaluated in *IARC Monographs* Volume 100 (IARC, 2012a, b, c, d, e, f), there was persuasive evidence of carcinogenicity in experimental systems. Since then, 150 environmentally relevant agents have been evaluated (or re-evaluated) in the *IARC Monographs* programme. Fig. 1.5 shows their classifications (moving outwards from Group 1 in the centre to Group 3 at the periphery), grouped by agent type and coloured by the evidence stream contributing to the evaluation. Notably, there have been contributions from several evidence streams for nearly all Group 1 agents. It is quite rare for human cancer evidence (either *sufficient* or *limited*) to form the sole basis for an evaluation (one example is radiofrequency electromagnetic field [RF-EMF] radiation). In Group 2A, there are numerous instances of *limited* human cancer evidence combined with either mechanistic or bioassay evidence. It is worth noting that nearly all these evaluations were based on occupational cancer epidemiology studies; this may be due to the generally higher exposure contrasts and well-characterized exposure information (leading to enhanced informativeness) in occupational settings (Loomis et al., 2018).

1.6 Conclusion

Cancer epidemiology studies have formed a crucial part of the evidence base for hazard identification since the early 1970s. Observational studies in which bias and confounding have been reasonably ruled out have been the main source of *sufficient* evidence leading to a determination that an agent is *carcinogenic to humans* – a process that has proven relatively conservative over the decades. The Preamble to the *IARC Monographs* calls for explicit examination of the potential for sources of bias (including confounding) to explain observed findings. This chapter provides examples of how such biases have been considered in recent *IARC Monographs* evaluations for agents found to have *limited* evidence (RF-EMF radiation, night shift work, and consumption of red meat) or *sufficient* evidence (opium consumption) of carcinogenicity in humans. Subsequent chapters explain concepts for explicitly evaluating the roles of confounding, information bias, and selection bias using these agents as examples, and demonstrate how these concepts may be incorporated into evidence synthesis.

Fig. 1.5. Environmentally relevant agents classified by the *IARC Monographs* in Volumes 101–136.



References

Baan RA, Straif K (2022). The Monographs Programme of the International Agency for Research on Cancer. A brief history of its Preamble. ALTEX. 39(3):443–50. doi:10.14573/altex.2004081 PMID:34164695

Blair A, Saracci R, Vineis P, Cocco P, Forastiere F, Grandjean P, et al. (2009). Epidemiology, public health, and the rhetoric of false positives. Environ Health Perspect. 117(12):1809–13. doi:10.1289/ehp.0901194 PMID:20049197

Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ (2008). False-positive results in cancer epidemiology: a plea for epistemological modesty. J Natl Cancer Inst. 100(14):988–95. doi:10.1093/jnci/djn191 PMID:18612135

Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, et al. (2011). Preventable exposures associated with human cancers. J Natl Cancer Inst. 103(24):1827–39. doi:10.1093/jnci/djr483 PMID:22158127

Cogliano VJ, Baan RA, Straif K, Grosse Y, Secretan MB, El Ghissassi F, et al. (2004). The science and practice of carcinogen identification and evaluation. Environ Health Perspect. 112(13):1269–74. doi:10.1289/ehp.6950 PMID:15345338

Cordina-Duverger E, Menegaux F, Popa A, Rabstein S, Harth V, Pesch B, et al. (2018). Night shift work and breast cancer: a pooled analysis of population-based case-control studies with complete work history. Eur J Epidemiol. 33(4):369–79. doi:10.1007/s10654-018-0368-x PMID:29464445

Health Council of the Netherlands (2012). Guideline for the calculation of occupational cancer risk values (Publication No. 2012/16E). The Hague, The Netherlands: Health Council of the Netherlands. Available from: https://www.healthcouncil.nl/binaries/healthcouncil/documenten/advisory-reports/2012/10/26/guideline-for-the-calculation-of-occupational-cancer-risk-values/advisory-report-guideline-for-the-calculation-of-occupational-cancer-risk-values.pdf.

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300. doi:10.1177/003591576505800503 PMID:14283879

IARC (1983). Miscellaneous pesticides. IARC Monogr Eval Carcinog Risk Chem Hum. 30:1–424. Available from: https://publications.iarc.who.int/48 PMID:6578175

IARC (1987). Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr Eval Carcinog Risks Hum Suppl. 7:1–440. Available from: https://publications.iarc.who.int/139 PMID:3482203

IARC (1994). Some industrial chemicals. IARC Monogr Eval Carcinog Risks Hum. 60:1–560. Available from: https://publications.iarc.who.int/78 PMID:7869568

IARC (1998). Report of an ad-hoc *IARC Monographs* Advisory Group on physical agents (IARC Internal Report 98/002). Available from: https://monographs.iarc.who.int/wp-content/uploads/2018/06/98-002.pdf.

IARC (2000). Ionizing radiation, Part 1: X- and gamma (γ)-radiation and neutrons. IARC Monogr Eval Carcinog Risks Hum. 75:1–491. Available from: https://publications.iarc.who.int/93 PMID:11203346

IARC (2010). Painting, firefighting, and shiftwork. IARC Monogr Eval Carcinog Risks Hum. 98:1–804. Available from: https://publications.iarc.who.int/116 PMID:21381544

IARC (2012a). Pharmaceuticals. IARC Monogr Eval Carcinog Risks Hum. 100A:1–435. Available from: https://publications.iarc.who.int/118 PMID:23189749

IARC (2012b). Biological agents. IARC Monogr Eval Carcinog Risks Hum. 100B:1–475. Available from: https://publications.iarc.who.int/119 PMID:23189750

IARC (2012c). Arsenic, metals, fibres, and dusts. IARC Monogr Eval Carcinog Risks Hum. 100C:1–501. Available from: https://publications.iarc.who.int/120 PMID:23189751

IARC (2012d). Radiation. IARC Monogr Eval Carcinog Risks Hum. 100D:1–341. Available from: https://publications.iarc.who.int/121 PMID:23189752

IARC (2012e). Personal habits and indoor combustions. IARC Monogr Eval Carcinog Risks Hum. 100E:1–575. Available from: https://publications.iarc.who.int/122 PMID:23193840

IARC (2012f). Chemical agents and related occupations. IARC Monogr Eval Carcinog Risks Hum. 100F:1–599. Available from: https://publications.iarc.who.int/123 PMID:23189753

IARC (2013). Non-ionizing radiation, Part 2: Radiofrequency electromagnetic fields. IARC Monogr Eval Carcinog Risks Hum. 102:1–462. Available from: https://publications.iarc.who.int/126 PMID:24772662

IARC (2018a). Drinking coffee, mate, and very hot beverages. IARC Monogr Eval Carcinog Risks Hum. 116:1–499. Available from: https://publications.iarc.who.int/566 PMID:31310458

IARC (2018b). Red meat and processed meat. IARC Monogr Eval Carcinog Risks Hum. 114:1–506. Available from: https://publications.iarc.who.int/564 PMID:29949327

IARC (2019). Preamble to the *IARC Monographs* (amended January 2019). Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

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

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

IARC (2022). 1,1,1-Trichloroethane and four other industrial chemicals. IARC Monogr Identif Carcinog Hazards Hum. 130:1–368. Available from: https://publications.iarc.who.int/611 PMID:37844156

Ioannidis JPA (2005). Why most published research findings are false. PLoS Med. 2(8):e124. doi:10.1371/journal.pmed.0020124 PMID:16060722

Lawlor DA, Tilling K, Davey Smith G (2016). Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866–86. doi:10.1093/ije/dyw314 PMID:28108528

Lo B, Field MJ, editors (2009). Conflict of interest in medical research, education, and practice. 1st ed. Washington (DC), USA: The National Academies Press. doi:10.1007/978-3-030-82673-4_6

Loomis D, Guha N, Hall AL, Straif K (2018). Identifying occupational carcinogens: an update from the IARC Monographs. Occup Environ Med. 75(8):593–603. doi:10.1136/oemed-2017-104944 PMID:29769352

Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L (2017). Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2(2):MR000033. doi:10.1002/14651858.MR000033.pub3 PMID:28207928

Lunn RM, Mehta SS, Jahnke GD, Wang A, Wolfe MS, Berridge BR (2022). Cancer hazard evaluations for contemporary needs: highlights from new National Toxicology Program evaluations and methodological advancements. J Natl Cancer Inst. 114(11):1441–8. doi:10.1093/jnci/djac164 PMID:36029241

Mandrioli D, Kearns CE, Bero LA (2016). Relationship between research outcomes and risk of bias, study sponsorship, and author financial conflicts of interest in reviews of the effects of artificially sweetened beverages on weight outcomes: a systematic review of reviews. PLoS One. 11(9):e0162198. doi:10.1371/journal.pone.0162198 PMID:27606602

McCullough LE, Maliniak ML, Amin AB, Baker JM, Baliashvili D, Barberio J, et al. (2022). Epidemiology beyond its limits. Sci Adv. 8(23):eabn3328. doi:10.1126/sciadv.abn3328 PMID:35675391

Michaels D (2008). Doubt is their product: how industry's assault on science threatens your health. Oxford, UK: Oxford University Press.

Pearce N, Blair A, Vineis P, Ahrens W, Andersen A, Anto JM, et al. (2015). IARC Monographs: 40 years of evaluating carcinogenic hazards to humans. Environ Health Perspect. 123(6):507–14. doi:10.1289/ehp.1409149 PMID:25712798

Samet JM, Berrington de Gonzalez A, Lunn RM, Schubauer-Berigan MK (2022). Commentary: Role and communications of cancer hazard determinations. Carcinogenesis. 43(2):79–81. doi:10.1093/carcin/bgac001 PMID:35016221

Samet JM, Chiu WA, Cogliano V, Jinot J, Kriebel D, Lunn RM, et al. (2020). The IARC Monographs: updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst. 112(1):30–7. doi:10.1093/jnci/djz169 PMID:31498409

Saracci R (2017). The hazards of hazard identification in environmental epidemiology. Environ Health. 16(1):85. doi:10.1186/s12940-017-0296-3 PMID:28793913

Savitz DA, Wise LA, Bond JC, Hatch EE, Ncube CN, Wesselink AK, et al. (2024). Responding to reviewers and editors about statistical significance testing. Ann Intern Med. 177(3):385–6. doi:10.7326/M23-2430 PMID:38373303

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 71(3):209–49. doi:10.3322/caac.21660 PMID:33538338

Taubes G (1995). Epidemiology faces its limits. Science. 269(5221):164–9. doi:10.1126/science.7618077 PMID:7618077

U.S. EPA (2022). ORD staff handbook for developing IRIS assessments (2022) (EPA/600/R-22/268). Washington (DC), USA: United States Environmental Protection Agency Office of Research and Development.

Wald NJ, Doll R, editors (1985). Interpretation of negative epidemiological evidence for carcinogenicity (IARC Scientific Publications No. 65). Lyon, France: International Agency for Research on Cancer. Available from: https://publications.iarc.who.int/208.

Zahm S, Bonde JP, Chiu WA, Hoppin J, Kanno J, Abdallah M, et al. (2024). Carcinogenicity of perfluorooctanoic acid and perfluorooctanesulfonic acid. Lancet Oncol. 25(1):16–7. doi:10.1016/S1470-2045(23)00622-8 PMID:38043561

Chapter 2. Causal diagrams to evaluate sources of bias

2.1 Introduction 24
2.2 Causal DAGs to evaluate sources of bias 24
2.3 Example: building a DAG for opium consumption and lung cancer 39
2.4 DAGs and specific sources of bias 41
2.5 DAGs and multiple sources of bias 51
2.6 Signed DAGs 53
2.7 Use of DAGs in evidence synthesis 56
2.8 Summary 57

CHAPTER 2.

Causal diagrams to evaluate sources of bias

Matthew Fox, Onyebuchi A. Arah, Sonja Swanson, and Vivian Viallon

2.1 Introduction

A key concern in studies of causal effects is identifying factors that prevent an observed association between an exposure and a disease from being equal to the true underlying causal effect of that exposure on the disease in the target population of interest. Although all analyses are subject to some systematic and random error, causal graphs, including causal directed acyclic graphs, can be used to attempt to understand which sources of bias may exist in studies and when sources of bias may prevent the identification of causation, such as the carcinogenic effect of an exposure.

Systematic error, or bias, occurs whenever the estimates generated in the study differ from the true causal effect for reasons other than random error. A key feature of systematic error is that, unlike random error, as the sample size in which the putative causal effect of interest is being studied increases, the systematic error is not expected to decrease. Epidemiologists generally focus on three main sources of systematic error: confounding, information bias, and selection bias, all defined in the Preface. These sources of bias are demonstrated in this chapter using causal diagrams and are discussed more extensively in the subsequent three chapters.

2.2 Causal DAGs to evaluate sources of bias

2.2.1 Introduction

Causal diagrams represent hypothesized relations between variables (Pearl and Mackenzie, 2018; Lipsky and Greenland, 2022). This section describes one type of causal graph, the directed acyclic graph (DAG), and discusses how to use DAGs to reason about bias, as well as ways in which they may be useful to *IARC Monographs* reviewers and to those evaluating research studies of causal effects. DAGs can also be used to identify a set of variables that is sufficient to control for confounding. Side Box 2.1 provides a brief history of causal diagrams; Side Box 2.2 gives their relation to the concepts of counterfactuals.

Why should those studying cause and effect learn about causal DAGs? And why specifically would an *IARC Monographs* Working Group reviewer want to learn to use DAGs? Because epidemiological studies are being used to inform public health policy decision-making, including specifically

Side Box 2.1. History of causal diagrams

Graphical methods have a long history in science; they can be traced back to Sewall Wright's path tracing approach (Wright, 1960) and to structural equation modelling, and were developed further by Glymour and Scheines (1986) and Pearl (2009). However, their use within epidemiology increased substantially after the publication of a seminal article by Greenland et al. (1999). They are related to but separate from counterfactuals (or counterfactual variables). A detailed explanation of counterfactuals is outside the scope of this chapter; however, a brief explanation of the relation between DAGs and counterfactuals is given in Side Box 2.2. (text continues on page 24)

Side Box 2.2. Relation between causal diagrams and counterfactuals

Causal DAGs are one formal language for causal inference, in which causal effects are defined in terms of counterfactual or potential outcomes. In brief, to understand the effect of a binary exposure X on a binary outcome Y, Y^x can be defined as the counterfactual outcome that would have occurred given exposure level x, and counterfactual contrasts of interest can be described as being about those counterfactuals (e.g. a causal risk ratio in a given population is $E[Y^{X=1}]/E[Y^{X=0}]$). Causal DAGs do not reference counterfactuals explicitly, because they encode the way in which data are realized (i.e. the data-generation process) rather than counterfactual worlds. The indirect link between causal DAGs and counterfactuals is that the absence of an arrow $X \rightarrow Y$ in a causal DAG encodes the sharp causal null that $Y^{X=1} = Y^{X=0} = Y$ for all individuals in the study (to put it simply, the exposure has no effect on the outcome for any individual investigated in the study). Pearl (2009) depicts the potential outcome as the outcome Y resulting from a mutilated DAG in which the arrow pointing into X is deleted and X is set to a specific value x depicting an intervention on X. Such mutilated or augmented DAGs are sometimes called post-intervention DAGs; they can be used to identify potential outcomes as the consequences of an intervention on an exposure X. Note that other types of causal diagram, including twin networks and single-world intervention graphs (SWIGs), have more explicit links to counterfactual theory. The focus here is on DAGs because of their ubiquity in practice, but it should be acknowledged that there are relative strengths and limitations to other formalizations of causal inference and causal graphs. (text continues on page 24)

in the context of *IARC Monographs* hazard identification, there is a need to communicate the findings of studies among researchers unambiguously, across the disciplines with which epidemiologists collaborate (e.g. toxicology and exposure science), and to stakeholders and decision-makers (Swanson, 2015). In practice, causal graphs facilitate communication between colleagues versed in causal graphs. Experience shows that disagreements within scientific teams over appropriate analyses often come down to the team members each assuming different causal structures that underlie the data in their minds. When these assumed structures are expressed as DAGs, they illuminate which questions are most important. For example, suppose the disagreement is over whether a particular covariate should or should not have been adjusted for in a study. When the competing DAGs are drawn, it may become obvious that the scientific consensus favours one graph over the other, thus ending the disagreement and clarifying which consensus evaluation (e.g. *sufficient*, *limited*, or *inadequate*, as described in Chapter 1) best describes the available evidence.

In cancer epidemiology, DAGs are used to summarize and formalize assumptions about the causal relations that may exist among variables relevant to the assessment of the carcinogenicity of the exposure under study. These DAGs represent our understanding of the data-generation process, meaning the set of variables, both measured and unmeasured, as well as the relations between them, that lead to the observed data we have to investigate for assessment of

a causal effect of an exposure on an outcome (Example 2.1a).

Causal diagrams are made of nodes (represented by the variables named in the diagram) and arrows. Each node represents one variable, and a single-headed arrow between two nodes represents an assumption of a possible causal effect between the corresponding two variables. A single-headed arrow is sometimes referred to as a directed edge, because the direction of the arrow is intended to indicate the direction of causation.

Technical details about the implications of the arrows are given in Side Box 2.3, but two details are noted here. First, arrows in DAGs can only be single-headed (i.e. directed). This means that there can be no feedback loops; therefore, bidirectional relations where two variables both seem to affect each other must be represented with time-dependent variables, which affect each other over time (examples are given in Section 2.2.4 and in Fig. 2.3 in Side Box 2.3, as well as in Section 3.2.4(a) and Example 7.6). Second, the graph must be acyclic, meaning that there is no place in the graph where one can start and trace a path following the direction of the arrows and get back to where one started. This is necessary because arrows encode time, given that for *A* to cause *B*, *A* must precede *B* (temporality). Thus, *A* cannot cause *B* in the future and have this, in turn, affect itself in the past. (If one thinks that both *A* and *B* can cause each other, this should be depicted in a DAG using several instances of *A* and *B*, indexed over time.) Satisfying the conditions of being directed (i.e. having only single-headed arrows) and being acyclic creates a DAG and allows for assessments of bias in published research and possible strategies to mitigate bias when designing and analysing studies (Example 2.1b).

As Example 2.1 illustrates, DAGs must be created by people using the best knowledge they have of the (causal) associations between the variables involved that lead to the observed data. Thus, DAGs do not by themselves indicate whether or not a variable is a confounder or a mediator; those creating the DAG must decide on what they believe to be the causal structure that created the data, and then use the rules of DAGs (see Section 2.2.4) to assess, under the assumptions encoded in the DAG, whether confounding or mediation (or some other of the structural relations described below) is present.

2.2.2 Paths

This section describes paths in DAGs and how they can be used to identify sources of bias. A path is defined as any sequence of consecutive arrows in the causal diagram, regardless of the directions of the arrows. In Fig. 2.1, examples of paths include red meat consumption → CRC, red meat consumption → BMI → CRC, and red meat consumption ← family history of CRC → CRC. Any path that always follows the direction of the arrows is called a directed path (e.g. red meat consumption → BMI → CRC), and any path that does not necessarily follow the direction of the arrows is called an undirected path (e.g. red meat consumption ← family history of CRC → CRC). Whereas the arrows represent causal relations between variables, the paths in the DAG can be used to identify whether we expect to see associations between any two variables; some of these associations may be causal, some of them may represent bias, and some of them will be a combination of the two.

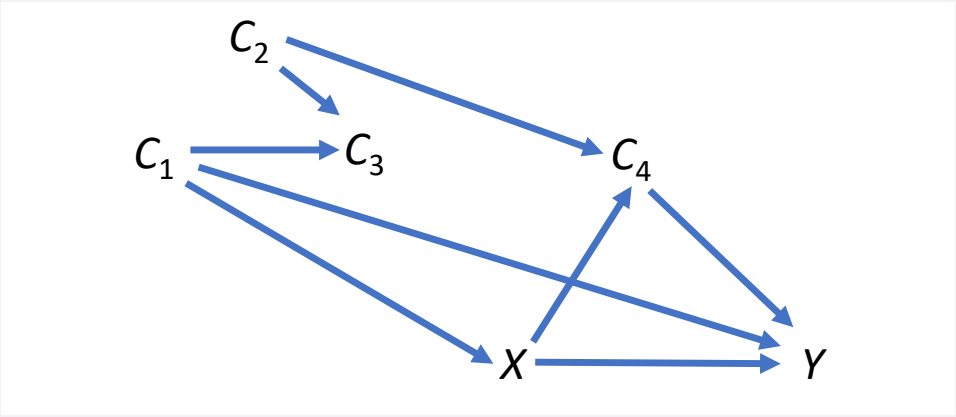
Example 2.1a. Motivation for creating a DAG for red meat consumption

For illustration, suppose that a team is reviewing the literature on whether red meat consumption is a hazard for colorectal cancer (CRC) and that there is a debate about whether it is critical for studies to have adjusted for family history of CRC and body mass index (BMI) to be considered high-quality evidence as part of the review (for simplicity, assume here that these are the only critical factors). Furthermore, suppose that some team members think that adjustments for both are necessary, whereas others think that only family history of CRC should be adjusted for and that adjusting for BMI may induce bias. A causal diagram, such as the hypothetical DAG of Fig. 2.1, could be drawn to help guide the group. (text continues above)

Side Box 2.3. Assumptions about arrows: causality and temporality

Assumptions in causal diagrams lie in the arrows that are absent from the diagram, as well as those that are present. The presence of an arrow from node C_1 to node C_3 in a causal diagram encodes an assumption that variable C_1 could be a (direct) cause of variable C_3, while the absence of an arrow from node C_3 to node C_1 stipulates the absence of a (direct) causal effect of variable C_3 on variable C_1. Arrows in DAGs encode assumptions about the existence of possible causal effects but not about the strength of such effects or their functional forms. The DAG in Fig. 2.2 indicates that Y could be influenced by C_1, C_2 (via C_4), C_4, and X, but not how much these variables may influence Y. In particular, the DAG does not reflect whether the effect of X on Y is assumed to depend on the levels of some combination of C_1, C_2, and C_4 (in such a situation, these would be effect modifiers).

Fig. 2.2. Illustrative example of a causal diagram for the study of a possible causal effect of exposure X on cancer risk Y.



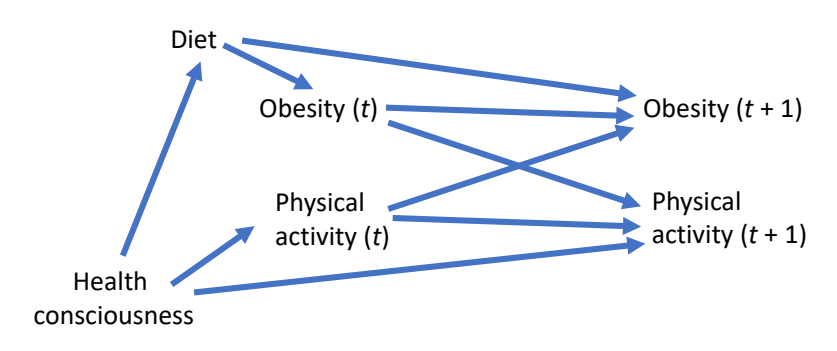
As noted by Hill (1965), temporality is a critical component of causality; for an agent to be causal, its presence must precede the development of the outcome. This implies that there can only be arrows $C_1 \rightarrow C_3$ and not also $C_3 \rightarrow C_1$ because causes (C_1) must precede their consequences (C_3). Cycles, or feedback loops, are usually prohibited in causal diagrams. If a directed path exists from C_1 to C_3, this implies that C_1 occurs before C_3; therefore, there cannot be another directed path from C_3 to C_1 because this would violate the temporality criterion. When no cycles are present in a causal diagram, it is said to be acyclic and is usually referred to as a DAG. To recap, the presence of an arrow from C_1 to C_3 in a DAG reflects an assumption that: (i) C_1 might have a direct causal effect on C_3; (ii) C_3 has no direct causal effect on C_1 (by the definition of a DAG, if there is an arrow from C_1 to C_3, there cannot be one from C_3 to C_1); and (iii) more generally, C_3 has no causal effect on C_1 (by the definition of a DAG, there cannot be any directed path from C_3 to C_1 if there is one from C_1 to C_3).

It is not always easy to determine the directionality of an arrow between two variables. Consider the example of obesity and physical activity. By increasing total energy expenditure, physical activity can help individuals to maintain their energy balance or even lose weight, so it can be inferred that lack of physical activity is probably a cause of obesity. However, excess weight also hampers physical activity, so that obesity can also be seen as a cause of lack of physical activity.

Side Box 2.3. Assumptions about arrows: causality and temporality (continued)

Such scenarios can be represented in DAGs by acknowledging the time-varying nature of exposures in the causal diagram (Fig. 2.3) and by drawing arrows (i) between (lack of) physical activity at any given time *t* and obesity at later times, *t* + 1, *t* + 2, … and (ii) between obesity at any given time *t* and (lack of) physical activity at later times, *t* + 1, *t* + 2, …. The expected association between obesity and (lack of) physical activity at any given time *t* can also be due to shared causes of these two variables; for example, health consciousness, although difficult to define and therefore rarely measured, may affect both (amount of) physical activity and obesity (e.g. through diet). In Fig. 2.3, it is assumed, for simplicity, that both diet and health consciousness are time-fixed, although time-varying versions could also be considered for these two variables. (text continues on page 26)

Fig. 2.3. Example of a longitudinal causal diagram, to illustrate a situation in which two variables affect each another but there is still no feedback loop.



Paths in the DAG represent key information for assessing bias in a study. If a DAG is a true representation of the data-generation mechanism, some paths create associations (whether causal or non-causal) between variables, while others do not. Therefore, it is crucial to specify the paths that comprise a DAG, especially those linking the exposure (here, red meat consumption) and the outcome (here, CRC), to identify whether any observed association (here, between red meat consumption and CRC) could only result from causation or may include bias. Note that, although this may seem counterintuitive, it is possible to enumerate paths that do not follow the direction of the arrows, and it will be seen later that there are good reasons to do so.

There are three basic path structures in causal diagrams: chains, forks, and colliders. These are each discussed next, along with their implications with respect to associations between two variables.

Chains and forks induce an association between the nodes at the opposite ends of the path, whereas colliders do not. Conditioning on (e.g. adjusting for) nodes lying within a path can change the observed associations between variables, depending on the structure type. These are each described in Table 2.1. Although this may seem an abstract discussion, these three structures can be used to help solve disagreements about which variables should be adjusted for to obtain a valid estimate of the causal effect of an exposure *X* on an outcome *Y*, and about which variables should not be adjusted for or controlled (Example 2.2).

Example 2.1b. Motivation for creating a DAG for red meat consumption (continued)

In Fig. 2.1, the creator of the DAG is representing a view that a family history of CRC affects both red meat consumption (e.g. having a family history of CRC might cause a person to consume less red meat) and risk of CRC (because genetic causes of CRC can be inherited). A more formal explanation of this is given later, but the DAG shows that family history of CRC is what we would typically think of as a confounder and that it would need to be adjusted for to validly estimate the causal effect of red meat consumption on CRC.

In Fig. 2.1, the creator of the DAG is also representing a view that red meat consumption can affect one's BMI and a finding that having a high BMI can cause CRC. This would be an illustration of a mediating pathway; part of the way in which red meat consumption causes CRC is by increasing one's BMI. Accordingly, adjusting for BMI in a statistical model would have the effect of removing some of the effect of red meat consumption on CRC from effect estimates and would lead to an inaccurate assessment of the true total causal effect of red meat consumption on CRC (again, this will be explained in more detail later). Thus, if this DAG is a correct representation of the way in which the data were generated, BMI is not a confounder but a mediator, and therefore should not be adjusted for analytically. As this example illustrates, DAGs can help groups, including *IARC Monographs* Working Groups, to clarify their thinking about how to infer causality and to communicate with each other about which variables should be adjusted for to determine whether something is a cancer hazard. Note that this example graph is somewhat simplified and may not be the consensus graph; if others were to draw another graph and justify their differences, then perhaps a different conclusion about adjustment for BMI or family history of CRC would be reached. (text continues on page 26)

Fig. 2.1. Illustrative example of a causal diagram for a study of a possible causal effect of red meat consumption on risk of colorectal cancer (CRC). BMI, body mass index.

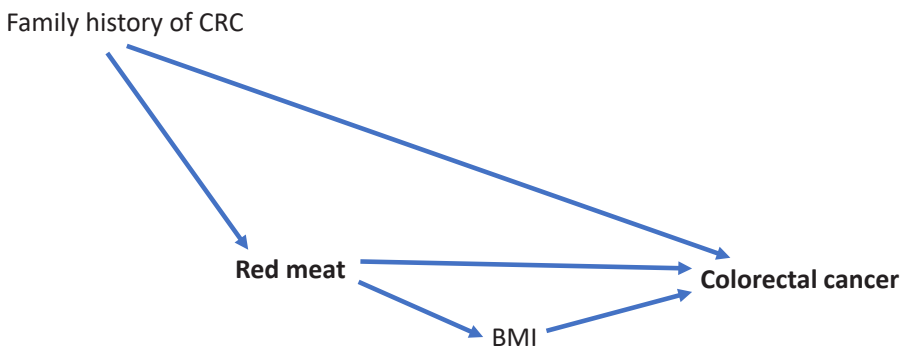


Table 2.1. Three basic structures in a directed acyclic graph and their implications for association and causation

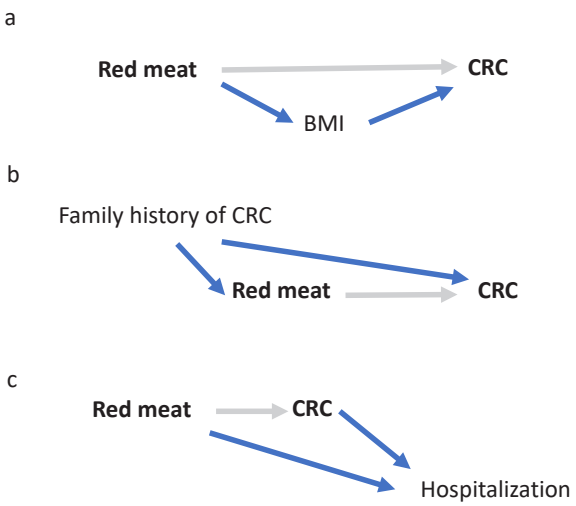
Name	Structure[a]	Example in Fig. 2.4	Explanation	Implications using example
Chain	$X \rightarrow Y$ $X \rightarrow M \rightarrow Y$	RM → BMI → CRC	A directed path in which all the arrows follow the same direction; the path from *X* to *Y* is open.	Creates a causal association between RM and CRC; BMI should not be adjusted for to estimate the (overall) causal effect of RM on CRC.
Fork	$X \leftarrow C \rightarrow Y$ $X \leftarrow Z \leftarrow C \rightarrow B \rightarrow Y$	RM ← FH → CRC	An undirected path in which there is a directed path from one node to two others (*C* to *X* and *C* to *Y*); the path from *X* to *Y* is open.	Creates a non-causal association between RM and CRC; path must be blocked (e.g. by adjusting for FH) to estimate the causal effect of RM on CRC.
Collider	$X \rightarrow S \leftarrow Y$ $X \leftarrow A \rightarrow S \leftarrow B \rightarrow Y$	RM → H ← CRC	An undirected path in which there are two directed paths from the outer nodes to a node in the centre (*X* to *S* and *Y* to *S*); the path from *X* to *Y* is blocked by collider *S*.	Creates no association between RM and CRC unless the collider is conditioned on (e.g. by adjusting for H). Controlling for H creates a non-causal association between RM and CRC (bias).

BMI, body mass index; CRC, colorectal cancer; FH, family history of colorectal cancer; H, hospitalization; RM, red meat consumption.
[a] *M* is a mediator of *X* and *Y*, *S* is selection, *X* is an exposure, and *Y* is an outcome; other variable letter names have no specific meaning and are used to illustrate the causal structure.

Example 2.2. Chains, forks, and colliders used in DAGs

The three basic structures used in causal diagrams are shown in Fig. 2.4, using red meat consumption as an example. (text continues on page 28)

Fig. 2.4. Three basic structures in a causal diagram illustrating a study of a possible causal effect of red meat consumption on risk of colorectal cancer (CRC): (a) chains; (b) forks; and (c) colliders. BMI, body mass index. Blue arrows indicate the structure being described; grey arrows are intended to make it easier to view the structure being illustrated.



(a) Chains

Directed paths are also called chains (Fig. 2.4a). In Fig. 2.1, the paths red meat consumption → CRC and red meat consumption → BMI → CRC generally imply an association between red meat consumption and CRC because in both cases red meat consumption may cause CRC, either directly (red meat consumption → CRC) or indirectly (red meat consumption → BMI → CRC). Chains represent causation; therefore, we are interested in the chains that follow pathways from the exposure to the outcome, whereas a chain from the outcome to the exposure would represent reverse causation (meaning a situation in which the variable that was assumed to be the independent variable was in fact the dependent variable, and vice versa). Because chains from the exposure to the outcome represent the causal effects whose effect sizes are to be estimated in causal epidemiological research, we do not want to disrupt these chains in a study design or data analysis.

One way to think about paths is as avenues for associations to flow along. Thus, if we are interested in the total effect of red meat consumption on CRC, conditioning on BMI (e.g. through restriction or covariate adjustment using methods such as stratification and regression) would create biased estimates in most situations in which the total effect is of interest, because this would block the causal path through BMI (i.e. stop the flow of the association from red meat consumption to CRC through BMI), hence eliminating part of the causal association. If the only way in which red meat consumption affected CRC was through changes in BMI, it would be expected that adjusting for BMI (or matching on it in the study design) would lead to an estimated null association between red meat consumption and CRC, when in fact there truly was a causal effect. Thus, if the hypothesized relations in the DAG are correct, reviewers would be wise to be concerned about a null result from a study of the effect of red meat consumption on CRC that adjusted for BMI, because it is possible that the reason the study showed a null result was not because there is no effect but rather because the authors inappropriately removed the effect by adjusting for BMI.

Key message

Although there may be circumstances in which we want to estimate the effect of an exposure that is not mediated through a specific pathway (in which case we might want to control for a variable on the causal pathway from exposure to outcome), we usually want to ensure that studies used for cancer hazard identification do not adjust for variables that lie on the causal chain from the exposure to the outcome.

There are two other basic structures in causal diagrams: forks (e.g. red meat consumption ← family history of CRC → CRC) and colliders (e.g. red meat consumption → hospitalization ← CRC). These correspond to simple forms of undirected paths.

(b) Forks

In a fork (Fig. 2.4b), there is (in a path) a node that has two arrows, each pointing to one other node. A path that contains only chains or forks in which no variables in the path are controlled (e.g. adjusted for statistically), except for the first and last variables in the path, is said to be open or unblocked. Non-causal associations between two variables (i.e. dependency) flow through forks; when a fork exists in a DAG, this implies that there is an association between the two variables at the end of the fork, even though that association is not causal. In Fig. 2.1, the path red meat consumption ← family history of CRC → CRC does not, on its own, imply any causal effect of red meat consumption on CRC, but it still typically induces a spurious association between the variables red meat consumption and CRC because the two have a common cause: family history of CRC.

As described in the Preface, confounding is the entanglement of a third factor (a confounder) in the association between an exposure of interest and an outcome of interest.

Key message

Open forks give rise to confounding (see Section 2.4.1 and Chapter 3) and represent confounding in DAGs.

Note that forks that indicate confounding can be made up of two chains, one going from a single variable to the exposure and one going from that same variable to the outcome. These paths can comprise a single arrow or can travel across several variables to get to the exposure and the outcome, as long as the direction of the arrows continues to lead from the node to the exposure (or the outcome). Given that forks represent biasing pathways,

they need to be closed (or blocked) to remove the bias. These are the confounding paths that researchers and readers of the epidemiological literature need to be concerned about when designing studies, analysing data, or conducting literature reviews, because if these open paths are not closed, they can create bias and may lead to the conclusion that there is a hazard when there is not, or vice versa, or they can cause overestimation or underestimation of the magnitude of the effect of an exposure on an outcome.

Seen another way, forks describe shared causes (of the variables at the arrowheads) that lead to confounding (Example 2.3a).

An open path can be closed or blocked by controlling for any inner node of that path that is not a collider (described next; see Fig. 2.4c) through adjustment, stratification, matching, regression, and so on (Example 2.3b). In DAGs, conditioning through analytical control of a variable is represented by drawing a box around that variable.

Note that, in this explanation of forks, confounders have not been discussed, only biasing pathways. This is because, although it may seem that the variable that is at the apex of the fork is the confounder, the confounding pathways can be blocked by controlling any variable on the pathway. Thus, removing confounding (sometimes called de-confounding) is much more important than identifying which variable is the confounder. Nonetheless, when DAGs are used, a variable is often called a confounder if it can be used (e.g. adjusted for) to block a confounding pathway.

(c) Colliders

A collider (Fig. 2.4c) is a node along a path with two arrows directly pointing to it along that path. Colliders do not, on their own, create a non-causal (biasing) pathway between the outer nodes; thus, a path that contains at least one collider that is not adjusted for is said to be blocked or closed (Example 2.4).

The bias induced by conditioning on a collider is not the most intuitive, but it can be understood by considering an idealized example, as given in Example 2.5.

A special but important case of collider stratification bias is when the collider is selection into the study. There are many reasons why people are enrolled (or choose to participate) in a study or drop out of a study. Because the study analysis can only be conducted among people who are enrolled in the study and for whom there is sufficient data, all studies are conditioned on selection. This means that if there is an effect to be estimated in a

Example 2.3a. Forks as depictions of shared causes in DAGs

The DAG in Fig. 2.1 indicates that red meat consumption and CRC have a shared cause, because one can trace a path following the arrows from family history of CRC to CRC and from family history of CRC to red meat consumption. Shared causes are typically thought of as confounding pathways that would need to be accounted for to find the causal effect of red meat consumption on CRC. (text continues above)

Example 2.3b. Conditioning or blocking of paths in DAGs

The association between red meat consumption and CRC, indicated as red meat consumption ← family history of CRC → CRC (the non-causal pathway that represents confounding), would be eliminated by conditioning on family history of CRC and could be partially removed by adjusting for any descendants of family history of CRC (i.e. variables with a directed path from family history of CRC to that variable), leaving only the associations indicated by the paths red meat consumption → BMI → CRC and red meat consumption → CRC, both of which are causal pathways. (text continues above)

Example 2.4. Depiction of colliders in DAGs

Suppose a group was reviewing a study in which data were collected on red meat consumption, hospitalization, and CRC and that red meat consumption increased the risk of being hospitalized (e.g. because of a heart attack), as did CRC. The path red meat consumption → hospitalization ← CRC does not create an association between red meat consumption and CRC, because this path is blocked by a collider. As long as the study design and analysis did not include conditioning on hospitalization (did not adjust for it, match on it, stratify on it, etc.), the results are likely to be valid (assuming that the DAG is correct and there are no other sources of bias). However, unlike forks, a blocked path can be unblocked by conditioning on one of the colliders (or any of its descendants, i.e. variables with an arrow towards that variable from the collider) (Berkson, 1946; Pearl, 2009). Although this is not necessarily intuitive, conditioning on colliders can induce spurious (non-causal) associations and result in collider stratification bias. Collider stratification bias is a bias that is created by conditioning on a collider, or an effect of (i.e. a descendant of) a collider. Thus, if the study adjusted for hospitalization, this would create bias in the association between red meat consumption and CRC and could make it appear that there was a hazard when there was not, or vice versa, or it could simply distort the magnitude of any real effect. Adjustment for hospitalization is elaborated on in Example 2.5, and colliders are described more intuitively in Side Box 2.4. (text continues on page 32)

population but the entire population, or a representative sample of that population, is not enrolled, it is only possible to estimate the effect in the selected sample; thus, the analysis is limited to a sample in which there are factors that lead to selection into the study. If both the exposure and the outcome are associated with selection into the study (either directly or indirectly through other forking paths), this can cause collider stratification bias. An example is provided in Section 2.4.3 (see also Chapter 5).

In conclusion, note again that for a diagram to be a DAG and therefore helpful for identifying and mitigating the impact of various sources of bias, it must be both directed and acyclic. A directed graph is one in which connections between variables must be drawn as single-headed arrows representing causality (an association cannot be implied between two variables using a dashed line without a specific cause, because any association must have a reason, as will be discussed later); furthermore, each causal path connecting more than two nodes sequentially in the DAG must contain arrows that point in the same direction. An acyclic graph is one for which there is no place in the diagram from which it is possible to trace a path following the direction of the arrows and get back to the starting point; in other words, no variable can cause itself in a DAG. Finally, for a DAG to be a causal DAG that can be used to identify sources of bias, the shared causes of any two variables in the DAG must also be represented. This means that a DAG that omits, for example, a common cause of any two variables is not a causal DAG, because the unknown causes of the two variables will not be mutually independent.

2.2.3 How to create a DAG

Researchers and reviewers often need to describe the data-generation process used in a study to assess the effect of an agent on cancer (i.e. the forces in the universe that create relations between variables, whether or not they are ever collected in a study, along with any relations created between variables in the process of study design and analysis) to decide which variables would ideally be controlled for to determine the causal effect in the study. Note that DAGs can also include several versions of a variable measured at different time points, as shown in Fig. 2.3 in Side Box 2.3. Drawing a DAG can help researchers and reviewers select such a set of variables. The drawing of DAGs requires expert knowledge of subject matter and of the data-generation process; teams researching causal relations or review panels determining the quality of existing evidence to ascribe causation (e.g.

Side Box 2.4. Collider bias

Consider the example where two binary {0,1} variables C_1 and C_2 have a common effect, a third binary variable C_3, as shown in Fig. 2.5. Conditioning on this common consequence C_3 usually creates a spurious association between C_1 and C_2, referred to as collider bias. For illustration, consider the situation where C_3 equals 1 if only one of C_1 and C_2 is equal to 1 but 0 if both are equal to 1 or neither is equal to 1, as shown in Table 2.2. If C_1 and C_2 are independent in the general population, then having information about C_1 for a person in the general population does not give us any information about the value of C_2 for that person. However, among individuals with $C_3 = 1$, if $C_1 = 0$ then necessarily $C_2 = 1$, and if $C_1 = 1$ then necessarily $C_2 = 0$. In other words, among individuals with $C_3 = 1$, having information about C_1 does give us information about C_2, highlighting that C_1 and C_2 are not independent after conditioning on C_3. In contrast, among individuals with $C_3 = 0$, if $C_1 = 0$ then necessarily $C_2 = 0$, and if $C_1 = 1$ then necessarily $C_2 = 1$. Therefore, within levels of C_3, C_1 and C_2 are perfectly inversely correlated. Conditioning on C_3 creates a spurious inverse association between C_1 and C_2. This is illustrated in Table 2.3 for a group of 400 individuals for whom there is no association in the total population between C_1 and C_2 but there is a perfect inverse correlation within C_3.

Fig. 2.5. Example of bias created by conditioning on a collider, C_3.

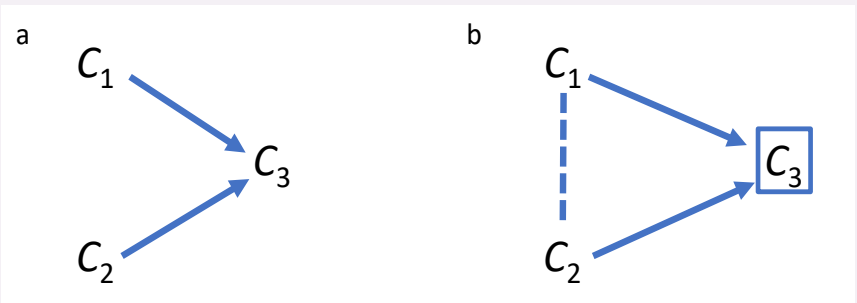


Table 2.2. Values of three variables, C_1, C_2, and C_3

C_1	C_2	C_3
0	0	0
0	1	1
1	0	1
1	1	0

Table 2.3. Frequency of cross-tabulation of C_1 and C_2, both overall and stratified by C_3

	Total			$C_3 = 1$			$C_3 = 0$	
	$C_1 = 1$	$C_1 = 0$		$C_1 = 1$	$C_1 = 0$		$C_1 = 1$	$C_1 = 0$
$C_2 = 1$	100	100	$C_2 = 1$	0	100	$C_2 = 1$	100	0
$C_2 = 0$	100	100	$C_2 = 0$	100	0	$C_2 = 0$	0	100
Total	200	200	Total	100	100	Total	100	100
% $C_2 = 1$	50%	50%	% $C_2 = 1$	0%	100%	% $C_2 = 1$	100%	0%

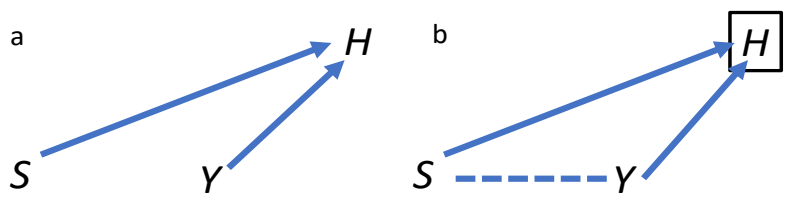
Side Box 2.4. Collider bias (continued)

Of course, collider bias is not restricted to binary variables. If smoking intensity (C_1) and alcohol intake (C_2) are both positively causally associated with the risk of a certain disease (C_3 = 1 if the individual develops the disease), then the association between smoking intensity and alcohol intake is typically less after conditioning on C_3, compared with that in the general population. This is because individuals with a low level of alcohol intake who develop the disease are more likely to have a higher smoking intensity than individuals in the general population. Similarly, individuals with a high level of alcohol intake who do not develop the disease are more likely not to smoke than individuals in the general population. (text continues on page 33)

Example 2.5. Example of a collider in a randomized controlled trial among downhill skiers

Suppose that we are interested in studying whether downhill skiing affects cancer risk, and in fact we even conducted a randomized trial in which participants were randomized to either never ski or ski frequently. In this randomized trial, we might expect to find no association between skiing and subsequent cancer risk, unconditionally. However, what if we restricted our analysis to only trial participants who went to the hospital sometime during the trial? Hospitalization might be a collider for a path skiing → hospitalization ← cancer, because both a cancer diagnosis and skiing accidents may lead to hospitalization (see the DAGs in Fig. 2.6). In an analysis in which only hospitalized participants were considered, skiing and cancer would probably be identified as inversely related: a person with a cancer diagnosis is less likely to be in the hospital for a skiing accident, and vice versa. Thus, we would not want to condition on hospitalization status in this trial, because it might make us wrongly conclude that skiing prevents cancer when, in this stylized example, it has no effect. (text continues on page 32)

Fig. 2.6. Example of a causal diagram depicting the relation between skiing (*S*) and cancer risk (*Y*) in a randomized trial: (a) full trial; (b) restriction to people who were hospitalized (*H*).



IARC Monographs Working Groups) typically possess such knowledge.

There are numerous approaches to creating a DAG. When creating a graph to support an analysis, it is critical to list all the variables that are considered essential in the data-generation process. For those reviewing the literature and therefore faced with the task of using the available studies (i.e. they are not conducting their own analyses or planning their own studies) to assess whether a hazard exists (but not necessarily how big the effect is), the approach could be simplified to focus on those factors that are likely to have the largest impact on creating bias. However, it is critical to note that this list should include all variables that might lead to a reasonable amount of bias, not only those measured in a study, because bias can exist even if the study did not account for it.

Key message

In identifying variables to be included in DAGs, priority could be given to those variables that might form part of an undirected forking path between the exposure and the outcome, because such undirected paths reflect potential sources of bias.

The next step is to link the variables with arrows, representing the possible causal relations between them, while remembering that the lack of an arrow between any two variables denotes a strong assumption about the absence of a causal relation between them. One approach to this would be to use one's best understanding of the causal relations to guide the first draft of a DAG. The DAG could then be presented to experts and stakeholders and revised based on feedback. Another approach would be to order all the nodes in time (with, say, left representing earlier time points and right representing later time points) and draw arrows from all variables that occur earlier in time to ones that occur later, only removing an arrow if there is a strong justification to do so based on expert knowledge that there is truly no causal effect of one variable on the other. Sometimes several competing DAGs must be considered when evidence- and knowledge-driven consensus remains elusive.

When the DAG describing the data-generation mechanism is complete, one can consider adding depictions of the study design and any bias that might have been created in the design process, focusing on selection bias and information bias (each described in the Preface and in detail later in this chapter). As noted in Section 2.2.2 and described further in Section 2.4.3, selection bias (collider stratification bias) can be introduced through the ways in which people are selected into or out of the study as well as into or out of analytical groups (through conditioning, matching, dropout, etc.). Selection can be represented as a node in the diagram (typically identified with the letter *S*), and the factors that are likely to cause participants to be in a study can be identified, whether they are factors determined by the design (e.g. selection of the study population in a case–control study, implementation of inclusion criteria) or factors that might determine the likelihood of participants self-selecting into a study (e.g. socioeconomic status [SES]).

Review panels assessing causation can use a DAG to determine whether the reviewers think that the analytical choices have removed all (or most) of the biases that existed. Furthermore, the DAG can be used to determine whether biases were created in the design or analysis that might prevent observation of a causal effect.

The process of drawing realistically complex DAGs can itself be complex, and some find it helpful to use software, such as DAGitty (https://dagitty.net/), which is freely available online, or the advanced Causal Fusion platform (https://www.causalfusion.net/), which is free but requires one to sign in. These tools can also provide an automated way to analyse a DAG for sets of variables, to control for confounding.

2.2.4 Rules of DAGs

(a) Causal paths

A review panel, such as an *IARC Monographs* Working Group, may wish to develop a DAG for an agent under evaluation and a type of cancer to help identify a reasonable set of variables to control for in the literature reviewed and another set that would ideally be ignored to give a valid result (Example 2.6a).

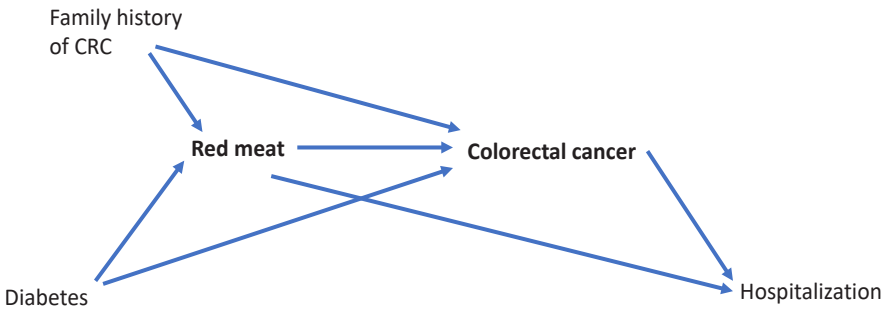
(b) Backdoor paths

In any study, but especially in observational studies where no randomization of the exposure occurred, it is necessary to consider that any observed association (e.g. between red meat consumption and CRC) may be a mix of any true causal effect and sources of bias (such as confounding by family history of CRC and diabetes). In DAGs, the most well known of these non-causal pathways are the open backdoor paths. These

Example 2.6a. A possible DAG for red meat consumption and colorectal cancer

The DAG in Fig. 2.7, which is a slightly more detailed version of the previous DAG for a study of the relation between red meat consumption and CRC, represents the DAG creator's understanding of the data-generation process. This DAG indicates that a family history of CRC is thought to affect red meat consumption (most probably by motivating one to consume less) and that a family history of CRC may also cause CRC. It also indicates that diabetes is thought to affect red meat consumption and that diabetes may cause CRC. Because there is no arrow from family history of CRC to diabetes or from diabetes to family history of CRC, the DAG also suggests that the two variables have no causal relation with each other and therefore would be expected to have no association, as revealed in the data. Note that there may also be an arrow from diabetes to hospitalization, but this is omitted for simplicity. (text continues on page 36)

Fig. 2.7. Directed acyclic graph for a study to assess the effect of red meat consumption on colorectal cancer (CRC).



are some of the paths that reviewers of epidemiological research want to identify to determine whether they are blocked, meaning either that they contain a collider or that a variable on the path has been controlled analytically. This is necessary to ensure that a causal effect has been satisfactorily assessed in a study, especially in a study for which the reviewer believes that the arrows drawn in the DAG represent strong effects. Backdoor paths are undirected paths (meaning that it is not necessary to follow the arrow directions) that can be traced from the exposure (e.g. red meat consumption) to the outcome (e.g. CRC) by tracing a path that begins with an arrow pointing towards the exposure (hence the term *backdoor path*) and ends with an arrow pointing towards the outcome. From the head of the arrow pointing towards the exposure (e.g. red meat consumption), the path can be traced in any direction to get to the outcome (e.g. CRC).

Although some backdoor paths indicate bias, not all do. For there to be a bias, the backdoor path must be open or unblocked so that the biasing associations can flow from a variable to both the exposure and the outcome (see Example 2.6b and Side Box 2.5). If such an open backdoor path can be traced, there will be a non-causal association between the exposure and the disease that is mixed with any causal effect of the exposure on the disease and that must be accounted for through some adjustment method to identify the true causal effect of an exposure *X* on an outcome *Y*. If the backdoor path is of the first type (e.g. containing forks only), the path can be blocked through analytical control of any variable between the exposure and the outcome. As noted, if the backdoor path is of the second type (e.g. containing a collider), it is blocked naturally only if one does not condition on the collider or a descendant of the collider, through methods such as statistical adjustment or design approaches such as restriction and matching. Otherwise, the path is open and must be closed again through the

Example 2.6b. Backdoor paths in a DAG for red meat consumption and colorectal cancer

In Fig. 2.7, there is an open (non-causal) backdoor path from red meat consumption to CRC, i.e. red meat consumption ← family history of CRC → CRC, which is one of the forking paths described previously. No variable along this path is a collider (there is no point in the path where one can enter a variable through the head of an arrow and also exit the same variable through the head of an arrow); if we have not controlled for any variables analytically, this path will create confounding, as noted earlier. This is because red meat consumption and CRC have a shared cause: family history of CRC.

Although one could try to identify all the shared causes of red meat consumption and CRC, the backdoor approach is a systematic way of identifying all the confounding pathways. This DAG shows that there is another unblocked backdoor path, as listed in Table 2.4. As well as red meat consumption ← family history of CRC → CRC, there is red meat consumption ← diabetes → CRC. Note that the path red meat consumption → hospitalization ← CRC is not a backdoor path, because although it does start at red meat consumption and end at CRC, it does not begin by going through an arrow towards red meat consumption. This path is also not an open (unblocked) path, because it contains a collider, hospitalization; thus, it is not a biasing pathway, as long as hospitalization is not conditioned on in the design or analysis. Seen another way, there is no variable in the path that one can start at and trace a path following the arrows and get to both red meat consumption and CRC. Thus, this path does not show a shared cause of red meat consumption and CRC, and hence there is no confounding. (text continues on page 37)

Table 2.4. All backdoor paths from red meat consumption to colorectal cancer in Fig. 2.7

Path	Backdoor?	Status	Reason for status	Path creates bias?
Red meat consumption ← family history of colorectal cancer → colorectal cancer	Yes	Open, unblocked	Fork with no collider, no variable on the path conditioned on	Yes, confounding
Red meat consumption ← diabetes → colorectal cancer	Yes	Open, unblocked	Fork with no collider, no variable on the path conditioned on	Yes, confounding
Red meat consumption → hospitalization ← colorectal cancer	No	Blocked	Path contains a collider	No, unless collider or its descendant is conditioned on

Side Box 2.5. Open backdoor paths

An open or unblocked backdoor path is a backdoor path that either

- does not contain a collider and no variable along it has been conditioned on; if the path contains a collider, it is naturally blocked, and there would be no variable along this path from which one could trace a path following the arrows to get to the exposure and another path following the arrows to get to the outcome (such a path would consist only of forks); or
- contains a collider (or a descendant of a collider) that has been conditioned on and otherwise does not condition on any non-colliders; this is because adjusting for a collider or a descendant of a collider opens the path that would have otherwise been blocked. (text continues on page 37)

analytical control (i.e. conditioning) of any non-collider between the exposure and the outcome.

For the purposes of evaluating the literature to see whether there is an effect of an exposure on an outcome, as an *IARC Monographs* Working Group might do, the DAG should represent all variables that are likely to lead to meaningful bias, even if they were not measured in the study being reviewed. For now, assume that all the variables in the DAG were measured. Now that each of the unblocked backdoor paths has been identified, it is necessary to assess whether the set of variables that were adjusted for is sufficient to control all the confounding (i.e. all the unblocked backdoor paths were blocked). The bias from a backdoor path can be removed by conditioning (through analytical control or design approaches) on any variable along the path. Thus, it is then necessary to identify a set of variables that will close (block) all the open (unblocked) backdoor paths. In the example of red meat consumption and CRC, the study could condition on family history of CRC and diabetes through analytical control, to block all the open backdoor paths (Example 2.6c), leading to an unbiased result.

It can be seen in Fig. 2.9 that all the existing biasing pathways in the DAG (i.e. all the unblocked backdoor paths) have been successfully blocked; however, in adjusting for the collider on the pathway red meat consumption → hospitalization ← CRC, a new biasing pathway is opened up: red meat consumption --- CRC. The true effect (red meat consumption → CRC) will now be mixed with the biasing pathway (red meat consumption --- CRC), giving a biased result. The resulting bias can be large, moderate, or small, depending on the context, including the strength of the associations and the distribution of the variables in the DAG. This example demonstrates that adjustment for variables in a statistical model can sometimes create rather than remove bias.

(c) Importance of bias

Before moving on to other examples, it is important to emphasize that DAGs can help to identify only whether a bias potentially exists, not its direction and magnitude. When using signed DAGs (see Section 2.6), it is sometimes possible to tell the direction of the bias; this is useful for identifying which biases can be ruled out as an explanation for an observed association (e.g. if the DAG identified a source of bias as operating downwards – a bias towards the null for a positive association – it could be concluded that an observed association is likely to be an underestimate). Where it is not possible to identify the direction or the magnitude of a bias, reviewers would need to consider using the various sensitivity analysis techniques presented in subsequent chapters before concluding that an identified source of bias would indeed be enough to change the interpretation of the study for the purposes of hazard identification.

Note also that DAGs cannot be used to solve every problem, and that there are some biases (particularly those involving the lack of a concept called faithfulness, which is beyond the scope of this chapter) that cannot easily be represented in DAGs. Another limitation of DAGs is that they do not readily depict interactions between variables.

2.3 Example: building a DAG for opium consumption and lung cancer

Suppose that an *IARC Monographs* Working Group comes together to evaluate whether opium consumption causes lung cancer, and that they are interested in using observational data to identify the hazard. When reviewing an observational study, such as the Golestan Cohort Study conducted in the Islamic Republic of Iran (Sheikh et al., 2020), the Working Group might be concerned that people who use opium are different from those who do not, with respect to factors that put a person at increased risk of developing lung cancer. A DAG could help the Working Group decide which variables should be controlled for in order for a study to be considered highly informative.

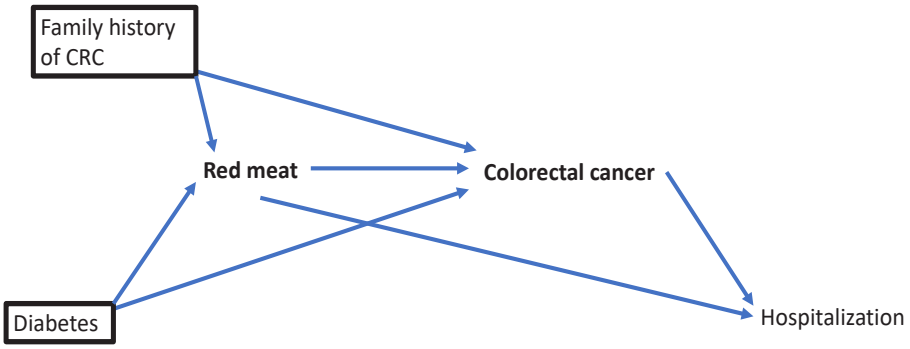
To begin to generate a DAG, the team would first draw the exposure and the outcome and then work through the shared causes of opium use and lung cancer as well as any other variables they think may be important in generating the data (Example 2.7a). The key is that expert knowledge is used to draw the DAG, not pure guesswork or the list of the variables that have been collected (Hernán et al., 2002).

It is critical to understand the issue hypothesized in Example 2.7a, where two variables (such as tobacco use and opium use) are associated through a third, possibly unmeasurable, latent variable (such as propensity to use substances). Having unmeasured latent factors in the DAG often creates difficulties in operationalizing what constitutes sufficient adjustment; however, such factors can still create substantial bias and

Example 2.6c. Conditioning to block backdoor paths in a DAG

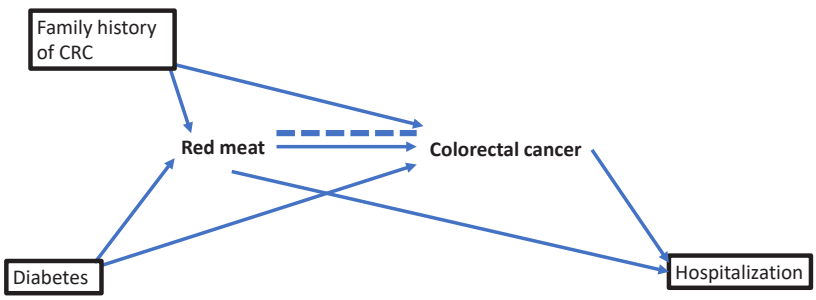
Fig. 2.8 shows that all the unblocked backdoor paths are now blocked.

Fig. 2.8. Directed acyclic graph representing the data-generation process for a study of red meat consumption and colorectal cancer (CRC) with additional conditioning on family history of CRC and diabetes.



Now, suppose that the results were adjusted for family history of CRC, diabetes, and hospitalization, because it was thought that hospitalization was a confounder. As noted in Table 2.4 for the path red meat consumption → hospitalization ← CRC, the only path from red meat consumption to CRC that goes through hospitalization is not an open path; thus, it creates no bias. Would the results still be a valid estimate of the effect of red meat consumption on CRC after conditioning on hospitalization? As discussed previously, conditioning on a collider (a variable with two arrowheads into it along a pathway) creates a non-causal association between the parents (i.e. the two variables that are causes of the collider), and this pathway creates bias. This is represented in Fig. 2.9 with a box around hospitalization (representing conditioning through statistical control) and a dashed arrow from red meat consumption to CRC (representing a non-causal association that has been induced between the two variables by controlling for the collider). As noted previously, spurious associations can be induced by adjusting for variables in the analysis; these are represented with dashed lines with no arrowhead. (text continues on page 39)

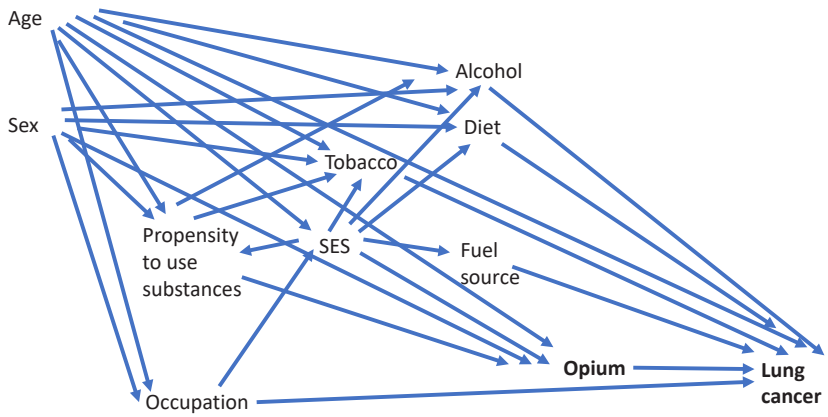
Fig. 2.9. Directed acyclic graph representing the data-generation process for red meat consumption and colorectal cancer (CRC) with additional conditioning on family history of CRC (confounder), diabetes (confounder), and hospitalization (collider), the last of which creates a non-causal association between red meat consumption and CRC. The dashed line represents an association created by conditioning on a collider.



Example 2.7a. Depicting shared causes in a DAG for opium consumption and lung cancer

It is known that tobacco use is a cause of lung cancer, and it is observed that people who use opium are more likely to smoke tobacco. However, it is not immediately clear how this association would occur. For a DAG to help identify a set of analytical variables needed to control for confounding, it is necessary to specify how associations occur in the data. Does tobacco smoking cause opium use? Does opium use cause tobacco smoking? Although, in a minimal number of individuals, either of these could be true, the review team believes that it is more likely that there is some shared cause that links the two. This shared cause could be a propensity to use substances, or something like SES. These causes are represented in the DAG shown in Fig. 2.10, because they are both considered to be likely sources of the association between opium use and tobacco use. (text continues on page 39)

Fig. 2.10. Directed acyclic graph for a study of the relation between opium use and lung cancer. SES, socioeconomic status.



therefore should not be omitted from the DAG.

For now, selection and measurement nodes, which are discussed later in this chapter, will be ignored and left out of the DAG. If it can be assumed that the DAG is correct, this now provides a model that can be used to identify a sufficient set of variables that need to be adjusted for to control for confounding (Example 2.7b). This process will be demonstrated in the discussion on confounding in this chapter. Note here that if the study investigators have not measured all the variables in the DAG (or have not measured them well), they may not have been able to remove all the confounding directly.

2.4 DAGs and specific sources of bias

2.4.1 Confounding

(a) Identification with DAGs

A confounder, which is a type of variable, can be distinguished from confounding, which is a bias that results from an unblocked backdoor path. A confounder is any variable that, when it has been controlled for, leads to a reduction in confounding. Given a DAG, researchers and reviewers can use simple rules to determine sets of variables whose control is sufficient to eliminate confounding bias (assuming that the variables are well measured). A set of variables is sufficient to eliminate all the confounding if (i) it blocks all open backdoor paths from an exposure *X* to an outcome *Y*, including any paths that may be opened through adjustment, and (ii) it comprises no descendant of the exposure *X* (i.e. no variable directly or indirectly influenced by the exposure) (Pearl, 2009).

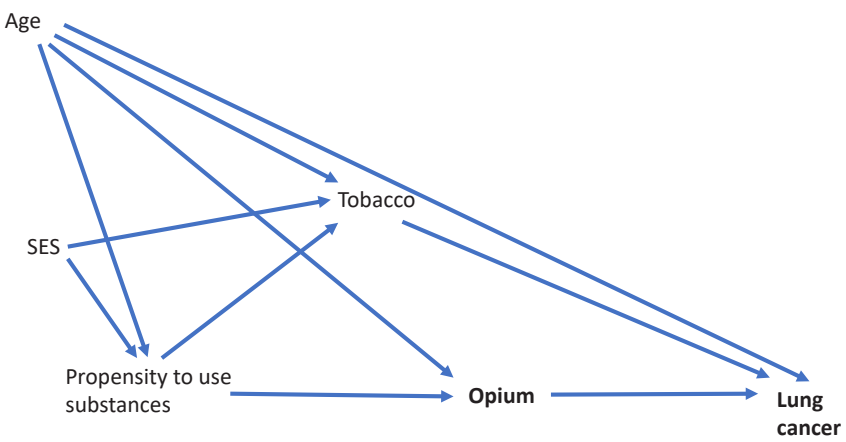
Previous definitions of a confounder used statistical terminology (e.g. a confounder is a variable that is

Example 2.7b. Simplifying a DAG for opium consumption and lung cancer to account for confounding

Fig. 2.10 is an example of a DAG that might have been created by an *IARC Monographs* Working Group. It is again necessary to reiterate that the DAG represents the full data-generation process (i.e. the set of variables and relations between them that led to the data observed in this study), not only the variables that were measured in the study. Note also that the data-generation process for a specific exposure–outcome pair might vary across populations. For example, in some countries occupational exposures to a specific carcinogen may be prevalent, and in others they may not, so the variable would not be included.

However, suppose that the Working Group members discuss the DAG and decide that only a few of the variables have effects (represented by the arrows) that are strong enough to represent substantial bias, based on their understanding of the strength of the effect the arrows represent. The DAG might then be simplified, as in Fig. 2.11. (text continues on page 41)

Fig. 2.11. Simplified directed acyclic graph describing the data-generation process for the relation between opium use and lung cancer, focused on variables that are thought to be likely to cause substantial bias. SES, socioeconomic status.



associated with both the exposure and the outcome); essentially, the definition of a confounder needs causal wording to properly distinguish it from other concepts, such as mediators (which are also variables that are associated with both the exposure and the outcome). Thus, a confounder has also been defined as any member of a minimally sufficient set of variables used for confounding control, such that dropping the variable from the sufficient set would lead to uncontrolled confounding (VanderWeele and Shpitser, 2013). Variables that are shared causes of both the exposure and the outcome qualify as confounders. Example 2.7c should make the distinction between confounders and confounding clearer.

Note, again, that free online software, such as DAGitty, can be useful here to identify all the sets of variables that would suffice to remove confounding through adjustment.

Key message

Confounding is the bias that is created by an unblocked backdoor path. Often, different sets of variables can be used to remove the confounding; which variables are identified as the confounders depends on which are to be used to remove the confounding.

Example 2.7c. Confounding by backdoor paths in the DAG for opium consumption and lung cancer

As depicted in Fig. 2.11, age is a confounder of the relation between opium use and lung cancer, because it has a causal effect on both opium use (likely reducing use) and lung cancer (increasing risk). Using the DAG terminology introduced in Section 2.2.2, the path opium use ← age → lung cancer is open and may thus create a spurious association between opium use and lung cancer. In plain words, and considering the example where both opium use (yes or no) and age (old or young) are binary variables for simplicity, people who use opium are more likely to be older, so that even in the absence of a causal effect of opium use on lung cancer, an association between them is expected because of the causal effect of age on lung cancer. Of course, how much bias this creates will depend on how strongly age affects both opium use and lung cancer.

As explained in Section 2.2.2, proper control for age (e.g. by stratification, matching, or adjustment) would block the path opium use ← age → lung cancer and remove the corresponding confounding bias. However, Fig. 2.11 also shows that more-complex confounding structures can exist. For example, it might be thought that people who use opium are more likely to smoke tobacco than people who do not use opium, probably through propensity to use substances. This makes propensity to use substances a shared cause (a forking path) of both opium use and tobacco use. Because tobacco use is known to cause lung cancer, there is now an unblocked backdoor path (i.e. a confounding path): opium use ← propensity to use substances → tobacco use → lung cancer. This path can be blocked by adjusting for propensity to use substances, but this is difficult to measure. Because the path can be blocked by controlling for any variable on it, adjusting for tobacco use would also suffice to remove confounding that works through this pathway. Therefore, valid studies of the association between opium use and lung cancer would be expected to include adjustment for tobacco use. Table 2.5 lists a selection of the backdoor paths from opium use to lung cancer (but note that there are more).

Table 2.5. Backdoor paths from opium use to lung cancer in Fig. 2.11 and their relevance to the control of confounding

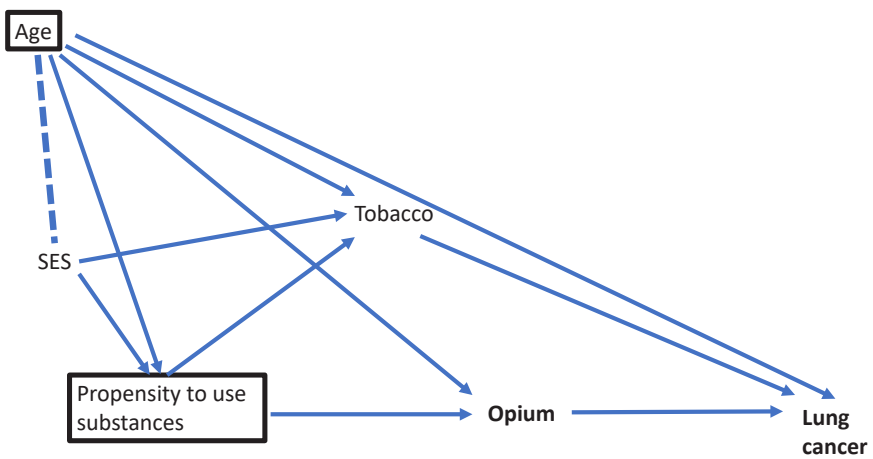
Number	Path	Status
1	opium use ← propensity to use substances ← SES → tobacco use → lung cancer	Open, unblocked
2	opium use ← propensity to use substances → tobacco use → lung cancer	Open, unblocked
3	opium use ← age → lung cancer	Open, unblocked
4	opium use ← age → tobacco use → lung cancer	Open, unblocked
5	opium use ← sex → tobacco use → lung cancer	Open, unblocked
6	opium use ← propensity to use substances ← age → tobacco use → lung cancer	Open, unblocked
7	opium use ← propensity to use substances ← age → lung cancer	Open, unblocked
8	opium use ← age → propensity to use substances → tobacco use → lung cancer	Open, unblocked
9	opium use ← sex → tobacco use ← age → lung cancer	Closed, blocked
10	opium use ← sex → tobacco use ← SES → propensity to use substances ← age → lung cancer	Closed, blocked
11	opium use ← propensity to use substances ← SES → tobacco use ← age → lung cancer	Closed, blocked
12	opium use ← propensity to use substances → tobacco use ← age → lung cancer	Closed, blocked

SES, socioeconomic status.

Example 2.7c. Confounding by backdoor paths in the DAG for opium consumption and lung cancer (continued)

In the example of Table 2.5, 12 backdoor paths are noted in a partial list of all the backdoor paths between opium use and lung cancer, of which only eight are open (1–8). The four that are blocked (9–12) are blocked because they contain a collider (e.g. sex → tobacco use ← age, sex → tobacco use ← SES, propensity to use substances ← SES → tobacco use, or propensity to use substances → tobacco use ← age). The eight unblocked paths all contain sex, age, or propensity to use substances; all of these unblocked paths could be blocked (i.e. removing all the confounding created by these forking paths) by controlling for these three variables. Although some of these variables are colliders on other paths and could thus open new paths, the new paths would all be blocked by one of the three variables in the set used to remove the confounding (Fig. 2.12). For example, propensity to use substances is a collider on the path SES → propensity to use substances ← age; if it were somehow possible to measure and adjust for propensity to use substances through stratification, regression, or some other method, a path would be opened between SES and age such that there would now be an open backdoor path: opium use ← propensity to use substances ← SES --- age → lung cancer. However, this path is already blocked by adjusting for propensity to use substances and sex; thus, no new bias is created. The key point here is that there may be instances where it is in fact necessary to control for a collider on a backdoor path from the exposure to the outcome to remove all the bias. This is fine as long as any new paths opened up by controlling for the collider are blocked.

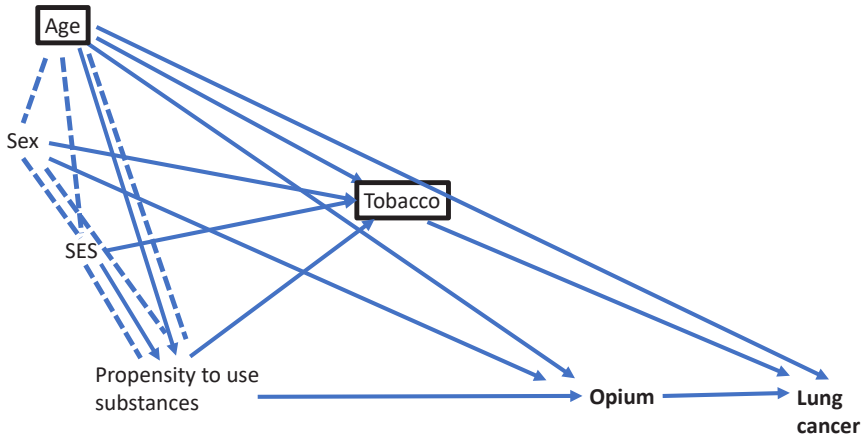
Fig. 2.12. Example of directed acyclic graph for a study of a possible causal effect of opium use on lung cancer, adjusted for age, sex, and propensity to use substances. SES, socioeconomic status. Dashed lines represent associations created by conditioning on a collider.



The eight unblocked paths also contain tobacco use or age (or both), so one could also adjust for both of these and block all the unblocked backdoor paths; however, note that tobacco use is a collider in each of the closed paths, so adjusting for tobacco use would open new pathways (Fig. 2.13). Nonetheless, although adjusting for tobacco use does open new pathways, none of them leads to a new unblocked backdoor path (confounding pathway) from opium use to lung cancer after adjusting for tobacco use and age, so this would also be an appropriate set. Because propensity to use substances may be difficult to measure in practice, adjustment for age and tobacco use may be an easier strategy.

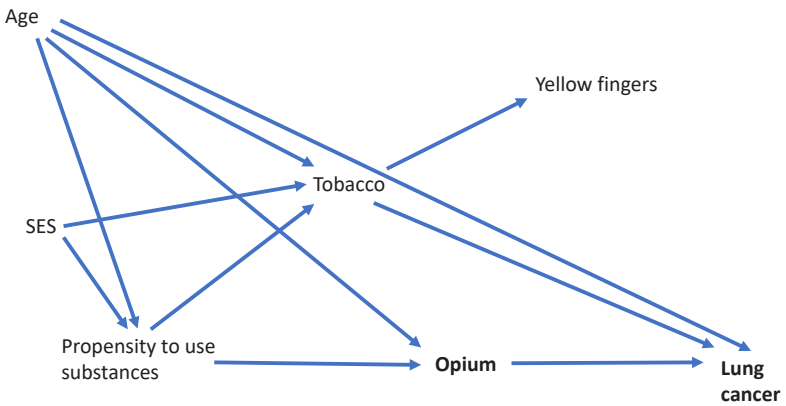
Example 2.7c. Confounding by backdoor paths in the DAG for opium consumption and lung cancer (continued)

Fig. 2.13. Example of a directed acyclic graph for a study of a possible causal effect of opium use on lung cancer, adjusted for tobacco use and age. SES, socioeconomic status. Dashed lines represent associations created by conditioning on a collider.



In Fig. 2.14, one variable is added: yellow fingers, which is an effect of tobacco use. The variable "yellow fingers" is a descendant of tobacco use (i.e. a variable affected by tobacco use), because it is possible to follow a directed path to get from tobacco use to yellow fingers. This demonstrates that adjustment for the descendant of a collider partially adjusts for the collider itself; in this instance, adjusting for yellow fingers (alone) would partially open a path between each of the parents. Although this may seem a silly example, it is meant to illustrate the approach used when investigators adjust for variables as a proxy solution in a situation where information on the variable they would have liked to adjust for (here, tobacco use) is missing. (text continues on page 42)

Fig. 2.14. Illustrative example of a directed acyclic graph for a study of a possible causal effect of opium use on lung cancer, as shown in Fig. 2.11, with an additional variable (yellow fingers). SES, socioeconomic status.



(b) Implications for study results

Key message

Given a DAG that describes the data-generation process under study, researchers and reviewers can determine whether a particular study sufficiently accounted for confounding bias. The rules given in this chapter can be used to check whether the analysis adjusted for a sufficient set of variables to remove all the confounding. If it becomes apparent that some confounding paths were not properly blocked, simple sensitivity analyses can be used to assess the magnitude of the residual confounding bias and whether this bias is likely to fully explain the observed association (see Chapters 3 and 6 for more details).

Sensitivity analyses can also be used to assess whether the size of the association between the exposure and the outcome would be larger than that observed had the bias been absent and, therefore, whether the data still suggest a cancer hazard. If the analysis involved a sufficient set of variables, confounding bias might still be present if some unobserved (or unobservable) variables were missing in the original DAG. Additional sensitivity analyses (Arah et al., 2008; VanderWeele and Arah, 2011; Arah, 2017), including analyses based on negative control exposures or outcomes (Flanders et al., 2022), can be carried out to explore this further (see Chapter 3).

2.4.2 Information bias

Another key source of bias that must be contended with in epidemiological research is information bias (Lash et al., 2021). As noted in the Preface, information bias results from the mismeasurement or misclassification of key variables. This section discusses ways of using DAGs to visualize different types of information bias when beginning to assess the possible impact that any mismeasurement of variables may have. This concept is discussed further in Chapter 4.

To obtain unbiased estimates of causal effects, accurate information is needed about the variables used in the study. Information bias occurs when the variables are not perfectly measured, and the mismeasured versions lead to a difference between the causal effect and the observed effect. For example, in the above-mentioned study of opium use and lung cancer, suppose that the study investigators assessed opium use with a questionnaire. Not all participants would provide accurate information about the amount of opium they typically used, for several reasons. Some may not accurately remember, and some may not want to tell the researchers, because opium use is usually illegal. Furthermore, if opium use was assessed after the lung cancer had already occurred, as may happen in a case–control study, it is possible that if the participants in the study thought there was a relation between opium use and lung cancer, the investigators may get more accurate information about those with lung cancer than those without; this could lead to a biased estimate of the true effect (often referred to as recall bias).

(a) Types of variables affected

All variables can be mismeasured to some degree. Although measurement error can be used as a catch-all term for mismeasured variables, mismeasurement that occurs in continuous variables is referred to as measurement error, whereas mismeasurement that occurs in categorical variables is referred to as misclassification. In both cases, it is possible to explore the impact of any potential bias created by the lack of perfect correspondence between the true value of a variable and its measured version. The next section first focuses on exposures and then discusses confounders.

(b) Identification with DAGs

Measurement error and misclassification can be depicted in DAGs, as demonstrated by Hernán and Cole (2009). With their approach, each factor in an analysis is represented with two variables: the true underlying variable and the measured version of that variable. Although the true version is almost never identified, a measurement approach is generally chosen that should be closely correlated with the actual values of the true variable to be measured.

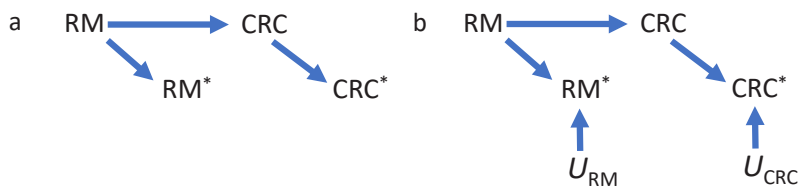
Example 2.8 describes the heuristic (Lash, 2007) that many researchers rely on when they note in their discussion sections that non-differential measurement error was likely to have biased their results towards no effect, despite the fact that this can be incorrect in a number of circumstances (van Smeden et al., 2020; Yland et al., 2022). In actuality, the structures can become more complex, and the direction of the bias can become unpredictable (at least in aggregate), as will be demonstrated. However, where the bias is probably towards the null, those who are simply trying to identify a non-null causal link between an exposure and

Example 2.8a. Depiction of non-differential measurement error in a DAG for red meat consumption and colorectal cancer

The DAG in Fig. 2.15a shows the presumed data-generation process for a study of the relation between red meat consumption and CRC. For simplicity, assume that there are no shared causes of the two variables. A measured version of each variable is also represented with the same variable name but with an asterisk * (indicating that this is the measured version). For each variable, the measured version is affected by the true variable; this creates an association between them. This is essential, because if there were no relation, there would be no reason to use the measured version.

Fig. 2.15b shows the same DAG, but the associated error terms are added, denoted by *U* with a subscript label related to the variable of interest; these explain the difference between the measured and true versions of the variable. Adding these error terms allows for the description of different types of measurement error, which can have different impacts on the results, and therefore on the inferences to be drawn from the study. In a study in which the authors estimate the effect of red meat consumption on CRC, what can in fact be estimated is the association between red meat consumption* and CRC*. Assuming that the measurement error in each does not depend on any other variable, in a very simple scenario, the association between red meat consumption* and CRC* might be expected to be attenuated compared with the true effect of red meat consumption in causing CRC, because red meat consumption* and CRC* are imperfect proxies for the true versions. (text continues on page 46)

Fig. 2.15. Directed acyclic graphs for a study of a possible causal effect of red meat consumption (RM) on risk of colorectal cancer (CRC): (a) representing the data-generation process, as well as measured versions of each variable (each represented with an asterisk); (b) with the addition of an associated error term (*U*) for each variable.



an outcome, and not the magnitude of the effect, might be able to focus less on this bias, because any observed association would probably have been stronger had the bias been absent.

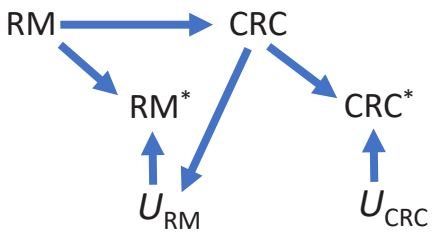
This measurement error can be classified as independent, because the error terms are independent of each other, and non-differential, because the error terms do not depend on the actual value of any other variable. Here, we focus on the distinction between differential and non-differential error and leave a discussion of dependent and independent error to Side Box 2.6. For an exposure, non-differential measurement error (using measurement error as a catch-all term here) typically means that the amount of measurement error in the exposure does not depend on the actual value of the outcome (although non-differentiality could be defined with respect to another key variable in the study). In many situations, the existence of non-differential error leads to the expectation of a bias towards the null. However, because DAGs do not imply anything about the magnitude of the effect of the arrows, it is not possible to say how much bias there will be; therefore, some may not find adding nodes for measurement to be beneficial. Quantitative bias analyses (Fox et al., 2021a) can be quite helpful in this situation, as discussed in Chapter 4.

When there is differential measurement error, there is no predictable direction of the bias; it can be

Example 2.8b. Depiction of differential measurement error in a DAG for red meat consumption and colorectal cancer

The DAG of Fig. 2.16 takes the DAG of Fig. 2.15b as a starting point but adds an arrow from CRC to U_{RM}. In this situation, the error in red meat consumption is affected by the true CRC status. This is an instance of differential measurement error, such as recall bias. For example, if the data on red meat consumption were collected before the cancer diagnosis, as would be depicted in the DAG of Fig. 2.15, it might be reasonable to assume that the error in information about red meat consumption is unrelated to whether a person develops CRC. However, if data on red meat consumption were collected by self-report after a diagnosis of CRC, the DAG in Fig. 2.16 might be more likely, because misreporting of red meat consumption might be different between those who did and did not have a diagnosis of CRC. This could occur in retrospective studies because those who have a diagnosis may spend more time trying to assess their exposures and may recall them more accurately than those who do not have a diagnosis. Alternatively, if people with a diagnosis believe that the cancer was caused by red meat consumption, they might overreport their red meat consumption compared with those who did not have a diagnosis (Lash et al., 2021). The key point with differential measurement error is that the error in one variable is related to the actual value of a second key variable (e.g. error in red meat consumption is related to actual CRC status). Thus, differential exposure measurement error typically means that the amount of measurement error in the exposure does depend on the actual value of the outcome (although it could be defined with respect to another key variable in the study). (text continues on page 46)

Fig. 2.16. Directed acyclic graph for a study of a possible causal effect of red meat consumption (RM) on risk of colorectal cancer (CRC), as well as measured versions of each variable (each represented with an asterisk) and the associated error term (U) for each variable, representing independent, differential measurement error.



towards or away from the null, and even our intuitions on the direction can sometimes be wrong (Greenland and Robins, 1985). Chapter 4 shows that having good information about the amount of measurement error in a variable stratified by any variable the error might depend on is the key to assessing the likely direction and magnitude of the bias.

Note that DAGs cannot indicate the magnitude of bias created by any information bias. See Chapter 4 for sensitivity analyses that enable the expert reviewer to consider whether information bias could meaningfully change causal conclusions from individual studies.

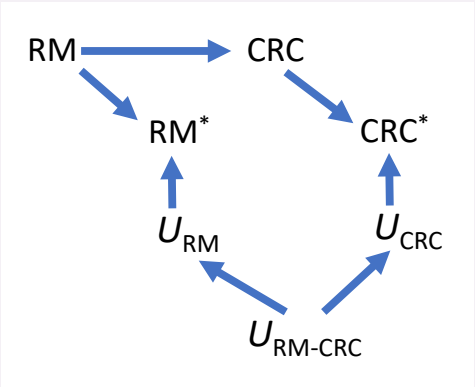
2.4.3 Selection bias

Although modern approaches to defining selection bias have focused on the use of causal diagrams, not all selection bias can be easily described using a DAG. Therefore, the focus here is on a mechanism by which selection bias occurs. As defined in the Preface, selection bias can occur when entry into or retention in a study is related to both the exposure and the outcome, although it should be noted that it can also occur because of the ways in which people are selected into analytical groups. In other words, selection biases create an association between exposure and outcome because of the way in which people are selected into or

Side Box 2.6. Dependent error

With respect to information bias, an informative scenario to consider is presented in Fig. 2.17; the error terms (unknown causes) are unrelated to the actual values of any other key analytical variables, but the error terms for both the exposure, U_{RM}, and the disease, U_{CRC}, have a shared cause, $U_{RM\text{-}CRC}$. This may occur if the same source (perhaps self-report) was used for both the exposure and the outcome. In such situations, the errors in the two variables are correlated, leading to non-differential but dependent measurement errors. This is also sometimes referred to as common-method or common-source bias (Podsakoff et al., 2003). In this situation, as was true in each of the previous DAGs (Fig. 2.15), because both red meat consumption and CRC were mismeasured, one would expect error in both variables. However, unlike in the previous two DAGs, the errors here are correlated with each other. This may be easiest to understand with a dichotomous exposure and a dichotomous outcome. If the red meat consumption (high or low) of 10% of study participants was incorrectly classified and the CRC status (yes or no) of 10% of study participants was incorrectly classified, then one would expect misclassification on both variables for 1% (the product of those two percentages) of study participants. However, in the DAG of Fig. 2.17, because the errors are correlated or dependent, one would expect misclassification on both variables for more than 1% of study participants. This is because if self-report was used for both the exposure and the outcome, people who are more likely to overreport their exposure might be more likely to overreport their outcome, and vice versa. In this scenario, as demonstrated in articles by Kristensen (1992) and Chavance and Dellatolas (1993), small amounts of non-differential but dependent measurement error can lead to strong bias away from the null for a truly null effect. Thus, to obtain valid estimates of the effect of an exposure on an outcome, it is critical to separate the sources for data on key variables in the study (Brennan et al., 2021). For example, if self-report was being used for red meat consumption, a medical record could be used to obtain information on the CRC diagnosis. Both could still be measured with error, but because the errors would not be correlated, the impact of the bias would often be smaller. Bias analyses are quite difficult to implement for dependent errors; therefore, *IARC Monographs* Working Groups should be cautious when reviewing studies that may contain dependent error.

Fig. 2.17. Directed acyclic graph representing the data-generation process for red meat consumption (RM) and colorectal cancer (CRC), as well as measured versions of each variable (each represented with an asterisk) and the associated error term (U) for each variable, representing dependent, non-differential measurement error.



Although this is not shown here, errors in measurement can also be both dependent and differential, creating a very unpredictable and potentially strong bias. In such situations, it can be nearly impossible to assess the true underlying causal effect of an exposure on an outcome. (text continues on page 47)

out of a study. In nearly all studies, except those that use a census of the study population such that there is no selection into the study, participants are selected into the study either by the investigators or by their self-selection into the study (or a combination of the two). Selection alone does not always lead to selection bias; it is only when the forces that lead people to be selected into or out of a study or the ways in which researchers select people into or out of analytical groups distort the true causal effect for the target population, leaving a biased association (see Chapter 5 for more information). With causal diagrams, it is easier to demonstrate when selection bias occurs. For now, note that an example of selection bias would be selection on or adjustment of a shared effect of the exposure and the outcome (i.e. a collider).

Key message

Selection bias can be produced at the time of study entry, at the time of sampling into a study, at the time of selection out of a study (e.g. loss to follow-up), or during analysis (analytical selection).

(a) Description and mechanisms

As is shown in more detail in Chapter 5, selection bias is a common issue across all study designs (Lash and Rothman, 2021). In randomized trials, there can be selection bias due to loss to follow-up. In cohort studies, selection bias can arise because of how the cohort is selected. Case–control studies can have selection bias due to inappropriate choice of control participants. This is not an exhaustive list but underscores the ubiquity of the problem. This section connects the commonality of these biases via DAGs.

(b) Depiction and identification with DAGs

Section 2.2.4 reviews how a closed backdoor path can be opened by conditioning on a collider (or a descendant of a collider) on that pathway. Such collider biases can occur from selection into the analytical dataset (Example 2.9).

Next, Example 2.10 elaborates on this simple causal DAG in the setting of a case–control study.

Loss to follow-up in any longitudinal study (e.g. randomized trials or cohort studies) can also create a selection bias, which can be depicted through conditioning on a collider in a causal diagram. In situations where loss to follow-up creates a bias, the time under observation in the study is related to the exposure and the outcome (Example 2.11).

Note that there are other ways in which loss to follow-up can be drawn in DAGs, but all of these structures reduce to the same issue: if we analyse only people who happened to continue to be observed in the study without further adjustment, we might be conditioning on a collider or a descendant of a collider in a path between exposure and outcome, as drawn in the DAG.

Examples 2.10 and 2.11 are only two ways in which causal graphs may depict selection bias; Chapter 5 describes others in detail. Let us now turn our attention to what can be done to avoid, address, or mitigate selection bias, and the role of DAGs in that process.

(c) Implications for study results

What can be done about selection biases? Returning to the DAGs in Fig. 2.20, there would be no biasing pathway if there were no box around *S*. But this, of course, is not usually a realistic situation and is beyond the control of someone trying to analyse or review existing data. However, when studies have loss to follow-up greater than some de minimis value (e.g. 5%), the approaches to evaluating the sensitivity of results described in Chapter 5 could be helpful.

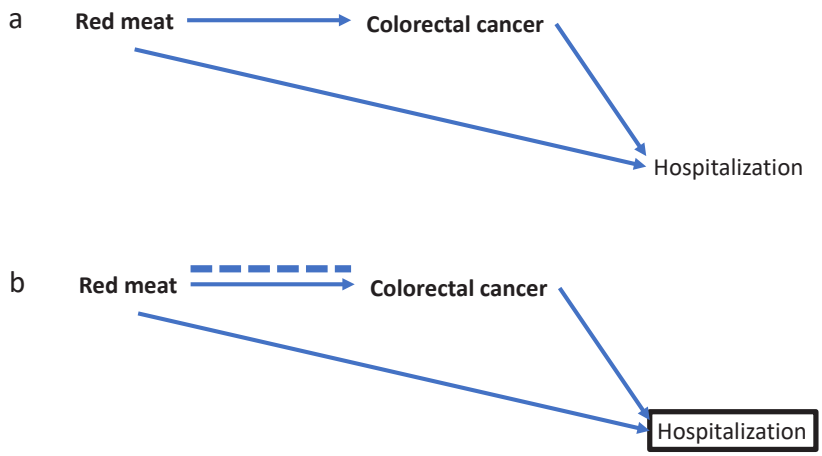
Selection bias can sometimes be minimized by design. For example, choosing control participants in a case–control study such that it is unlikely that a path exists between the proposed causal agent and the selection of the control group minimizes the bias created in the DAG in Fig. 2.19a, even if it cannot be guaranteed to prevent it completely. As another example, in studies where outcome assessments are obtained from routinely collected data rather than onerous study visits, loss to follow-up may be minimized through this reduced participant burden (although at a potential cost of information bias).

Even if selection bias has not been minimized by design, reviewers can use DAGs to ensure that in the studies being evaluated the analytical steps were taken to mitigate the bias to the best extent possible with the available data. The graph for loss to follow-up in Fig. 2.20b shows that the open pathway can be closed by adjusting for a variable on the newly opened path, namely SES. Chapter 5 describes in more detail the options for these adjustments, as well as ways to reason about the direction

Example 2.9. Depiction of collider bias (by hospitalization)

Suppose, for a study of the association between red meat consumption and CRC, that the DAG in Fig. 2.18a depicts the data-generation process but that in this situation the study being reviewed was conducted only among hospitalized patients. In other words, the study design conditioned on hospitalization, as shown in the DAG in Fig. 2.18b. Conditioning on hospitalization opens up the path red meat consumption → hospitalization ← CRC; therefore, there is an open path between red meat consumption and CRC other than the causal path of interest, and this new open path could explain any observed association between red meat consumption and hospitalization in the dataset. (text continues on page 50)

Fig. 2.18. Simple selection-bias diagrams showing (a) selection on hospitalization as a collider and (b) bias from conditioning on hospitalization. The dashed line represents an association created by conditioning on a collider.



and magnitude of bias when adjustment is not possible.

Key message

With respect to DAGs and review panels, perhaps the most useful implication of DAGs for selection bias is in identifying when selection bias is likely in a published study, and then using DAGs as a guide to inform a possible bias analysis (Fox et al., 2021b; Chapter 4) or sensitivity analysis for whatever remaining biases exist within the evidence at hand.

2.5 DAGs and multiple sources of bias

2.5.1 Identifying multiple sources of bias

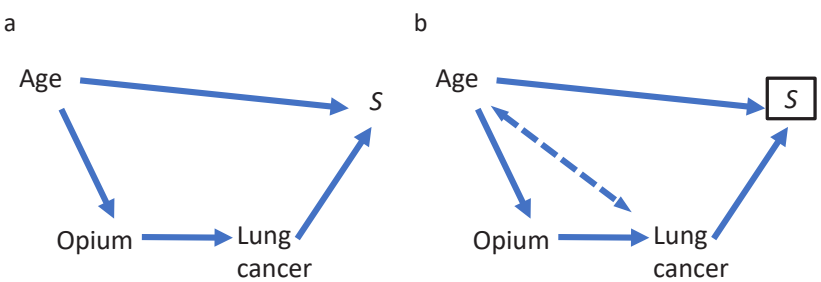
As noted previously, the full data-generation process can be represented in DAGs by including those sources of bias that occur in the population (e.g. confounding) and those that occur because of the study (e.g. selection bias and information bias); this will allow for a full picture of the ability of a study to identify causal effects from the observational data. When assessing the full impact of bias on study results, it may be necessary to think through how sources of bias interact with each other. It is not immediately clear from looking at a DAG whether two sources of bias will be additive in terms of their impact on study results or, if they act in opposite directions, whether they might cancel each other out (Greenland, 2005).

A simple scenario of how two sources of bias might interact with each other is information bias (in the measurement of exposures, outcomes, or other variables) and confounding. Suppose that a DAG representing the underlying data-generation process is used to identify a sufficient set of variables to control for all the confounding.

Example 2.10. Selection bias in a DAG for opium consumption and lung cancer

Consider an investigation of the effects of opium use on lung cancer that included a case–control study in which participants in the control group were selected from hospitalized patients. Let *S* denote an indicator of being included in the case–control study. In a case–control design, there is an arrow from the outcome to selection (lung cancer → selection [*S*]) by definition: having lung cancer (*Y* = 1) increases the probability of being selected into the study as a case participant (*S* = 1). Ideally, control participants are selected so that they represent the exposure distribution that gave rise to the case diseases; therefore, there should be no arrow from the exposure to selection (Fig. 2.19a). However, perhaps in this hypothetical study control participants were selected who had been hospitalized for other reasons, and people who were hospitalized were more likely to be older than the general population. In that situation, the DAG may look more like the DAG in Fig. 2.19b, where this choice of control participants creates a biasing pathway (lung cancer --- age → opium use). Side Box 2.7 describes selection bias in matched case–control studies. (text continues on page 50)

Fig. 2.19. Selection-bias diagram showing (a) selection (*S*) as a collider and (b) bias from conditioning on selection in a case–control study with control participants selected from among people with a condition related to the exposure. The dashed line represents an association created by conditioning on a collider.



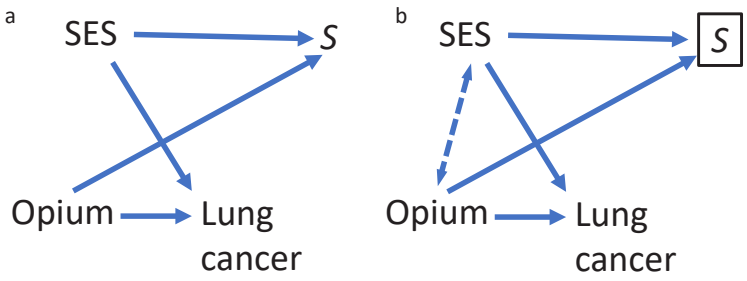
Side Box 2.7. Selection bias in matched case–control studies

Note that the type of biasing pathway shown in Example 2.10 also occurs in matched case–control studies in which confounders of exposure and outcome are chosen as matching factors. The matching creates a selection bias; this is why in such studies it is necessary to adjust for matched variables to remove the bias (i.e. block the backdoor path that is opened by the matching). Matching in a case–control study does not remove confounding, as is often thought; rather, it creates an efficient population within which to control for the confounding. This is sometimes referred to as selection bias by design. (text continues above)

Example 2.11. Depiction of collider bias (from loss to follow-up)

Consider a hypothetical cohort study of opium consumption, as shown in Fig. 2.20a, in which opium use is directly related to why participants continue to be observed in the study (i.e. were not lost to follow-up), and in which SES affects both cancer risk and the likelihood of remaining observed. This might occur if people who use opium are less likely to continue in a study than those who do not. Here again, there is a collider on the pathway opium use → selection ← SES → lung cancer. If the collider on this pathway were not conditioned on, it would not create any bias because this path is closed due to the collider. Because in this situation selection represents loss to follow-up, by definition selection has been conditioned on, because it is only possible to analyse people for whom there are data, and this gives the DAG in Fig. 2.20b, which now has an open unblocked backdoor path from opium use to lung cancer: opium use --- SES → lung cancer. (text continues on page 50)

Fig. 2.20. Simple selection-bias diagram showing (a) loss to follow-up as a collider (*S*) and (b) bias from conditioning on loss to follow-up. *S*, selection; SES, socioeconomic status. The dashed line represents an association created by conditioning on a collider.



However, suppose in addition that there is an imperfect measure of the key confounders. Although the DAG can show which variables are necessary to remove the impact of confounding, the mismeasurement of those variables can lead to imperfect control. If the mismeasurement is severe enough, the residual confounding will be quite strong. Approaches to evaluating the direction and magnitude of such residual confounding are described in Chapter 3.

2.5.2 Representing and identifying multiple sources of bias in a DAG

Representing the data-generation process to identify confounding, adding selection nodes to represent the selection of the study population and possibly nodes to represent selection out of the study (i.e. loss to follow-up), and adding nodes to represent the measured version of each variable and any biasing structures related to the error terms can create a very complex DAG. Although this process would ideally be followed for all variables, it may be helpful to focus on the variables that represent the largest sources of bias. However, this is challenging, because without knowing the impact that a particular source of bias has (say through a bias analysis method, described in later chapters), we are left with our intuition and our expert experience as to which biases are most important. It is recommended to start with as complete a DAG as possible for a particular study or set of studies and then remove biasing pathways that are thought to have minimal impact on the study results. See Chapter 6 for methods on this topic for triangulation.

2.6 Signed DAGs

The DAGs introduced thus far do not directly indicate the direction of a bias, but signed DAGs offer an approach that aids in identifying the direction of a bias. While signed DAGs can clarify many forms of bias (VanderWeele and Hernán, 2012), the focus here is on their use in understanding confounding. Suppose that an *IARC Monographs* Working Group is considering one uncontrolled

(dichotomous) confounder that does not modify the effect of the exposure on the outcome on the chosen effect measure scale (e.g. relative risk or risk difference) and wishes to understand whether this source of uncontrolled confounding is likely to explain all or some of the observed non-null association.

Signed DAGs are augmented to contain + or − symbols along the arrows to indicate the net or average direction of the effect (VanderWeele et al., 2008). A positive sign (+) indicates that an increase in (or the presence of) the variable at the tail of the arrow leads to an average increase (or no change) in the variable at the arrowhead, while a negative sign (−) indicates that an increase in (or the presence of) the variable at the tail of the arrow leads to an average decrease (or no change) in the variable at the arrowhead (see Side Box 2.8).

When one thinks about paths that can run between several variables and therefore have several arrows, rather than a path that is simply between two variables and has a single arrow, the sign of a path in a signed DAG is given by the product of the signs of its component arrows. An *IARC Monographs* Working Group that is interested in assessing the likely direction of confounding can begin by augmenting an existing DAG (as described in the previous sections) with these + or − symbols to represent the well-informed hypothesized direction of the relations.

To demonstrate how signed DAGs work, Example 2.12 extends the simple DAG shown in Fig. 2.4b. With two arrows and two possible signs that could be applied to the arrows, there are four possible scenarios and two possible results; a positive sign in the result describes the direction of the confounding as representing positive or upward bias (i.e. the bias leads to an observed estimate that is higher than the true effect), and a negative sign represents negative or downward bias (i.e. the bias leads to an observed estimate that is lower than the true effect). The net direction of the confounding created by each scenario follows the multiplication rules of positive and negative numbers, as shown in Table 2.6 and Fig. 2.23.

Side Box 2.8. Interpreting lack of change in signed DAGs

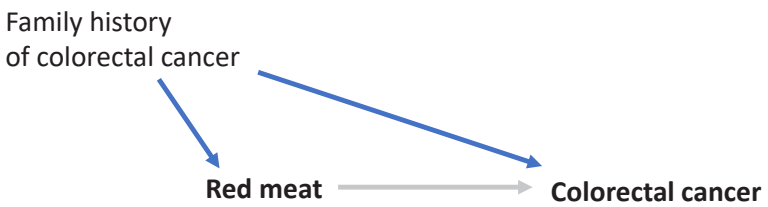
When interpreting signed DAGs, it may seem odd that "or no change" is included; it might be assumed that having no arrow would imply no change. This would be a reasonable assumption, but a lack of an arrow specifically implies no effect of the exposure on the outcome for any individual in the population (i.e. the sharp null). In contrast, there could be no average effect in the presence of an arrow if the number of people who experienced harmful effects was the same as the number of people who experienced preventive effects, such that the observed association averaged to the null. This might occur if the exposure prevented the outcome for some people in the population and caused it for other people, as might occur for seat belt use and death in an automobile accident. Although in this example it would be unlikely that the number of people for whom the exposure causes the outcome would be the same as the number of people for whom it prevents the outcome, in some exposure–outcome pairs such a result may be possible. In such a situation, on average, the exposure would be inferred to have no effect, even though for some people the exposure caused the outcome and for other people it prevented the outcome.

In this chapter, signs are only used in signed DAGs under the weak monotonicity assumption of non-decreasing (i.e. positive) or non-increasing (i.e. negative) average causal effects to assess the sign or direction of uncontrolled confounding due to an unmeasured confounder. Under this monotonicity assumption, a positive average monotonic effect, depicted as a positive sign on an arrow, means that increasing the value of the variable at the tail of the arrow always increases or leaves unchanged the average value of the variable at the arrowhead, for all values of the other covariates adjusted for in the analysis, in the entire population. Similarly, a negative average monotonic effect, depicted as a negative sign on an arrow in the DAG, means that increasing the value of the variable at the tail of the arrow always decreases or leaves unchanged the average value of the variable at the arrowhead, for all values of the other covariates adjusted for in the analysis, in the entire population. (text continues above)

Example 2.12a. Depiction of signed DAGs

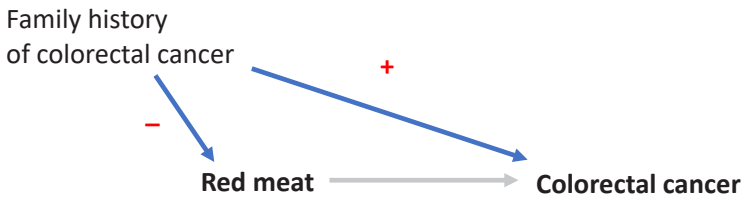
In this example, there is a concern that a family history of CRC might confound an estimate of the effect of red meat consumption on CRC, as in Fig. 2.21. Furthermore, suppose that the study results indicated an increased risk of CRC associated with red meat consumption. As explained earlier in this chapter, the unblocked backdoor path red meat consumption ← family history of CRC → CRC, which represents a source of confounding bias, would need to be addressed to determine the causal effect of red meat consumption on CRC. If the study being assessed did not control for family history of CRC, then before dismissing the study, the *IARC Monographs* Working Group would want to decide whether the uncontrolled confounding might explain the finding. In other words, the reviewers would want to know: if family history of CRC had been controlled for in the analysis of the study, is it at least possible that the true effect would have been null? Here, signed DAGs can help.

Fig. 2.21. Fork structure denoting confounding by family history of colorectal cancer in a study of a possible causal effect of red meat consumption on risk of colorectal cancer.



The first step in using a signed DAG is to hypothesize about the direction of the effect of the blue arrows. The arrow from red meat consumption to CRC has been left grey to indicate that this is the causal relation that is to be assessed. Fig. 2.22 shows the hypotheses about the blue arrows. On average, a family history of CRC is expected to increase the risk of developing CRC, perhaps due to a genetic predisposition; this is depicted with a positive sign to indicate a positive association. Furthermore, a family history of CRC is hypothesized, on average, to decrease red meat consumption, given the awareness of a potential link between the two and a desire of people with a family history of CRC to avoid developing the disease. This hypothesis of a negative association is depicted with a negative sign in the DAG. In this scenario, it is possible to identify the likely expected direction of this bias. (text continues on page 56)

Fig. 2.22. Signed DAG for assessing the direction of confounding by family history of colorectal cancer in a study of a possible causal effect of red meat consumption on risk of colorectal cancer.



Example 2.12b. Using signed DAGs to determine the possible impact of biases

Returning to the signed DAG in Fig. 2.22, because the arrow from family history of CRC to CRC is positive and the arrow from family history of CRC to red meat consumption is negative, the probable net bias in the association between red meat consumption and CRC in a study in which family history of CRC was not adjusted for would be downwards or negative. This means that if a positive association (e.g. relative risk [RR] = 1.6) was observed between red meat consumption and CRC, because the bias was likely to be downwards (towards the null), if there had been data on family history of CRC and it was adjusted for, the estimate of the effect would be expected to be even larger than what was observed (in this example, RR > 1.6). In other words, because the negative uncontrolled confounding from the signed DAG and the estimated positive association from the study have opposite signs, the observed association probably underestimated the unobserved effect adjusted for the unmeasured confounder. Accordingly, such a study could not be dismissed, given that the goal was to determine whether consumption of red meat is carcinogenic and not the magnitude of the effect (which would indeed be biased).

Suppose, however, that the *IARC Monographs* Working Group encounters a study in which the observed association was that red meat consumption was associated with a reduced risk of CRC (e.g. RR = 0.8, indicating a negative association), but the study also did not adjust for a family history of CRC. In this situation, the Working Group would make all the same assumptions as before, that family history of CRC increases risk of CRC but decreases red meat consumption, yielding negative uncontrolled confounding. However, because the observed association was negative (i.e. protective against cancer), the expected bias, which is also negative, could have been part of the observed association, and adjusting for the unmeasured family history of CRC could have removed some or all of the observed association between red meat consumption and CRC. Thus, the result would probably have been less negative (closer to the null, or even positive) than what was observed (in this example, RR > 0.8). In this scenario, strong conclusions cannot be drawn. The true unbiased result could have been a less protective, a null, or a harmful effect of red meat consumption, in which the negative uncontrolled confounding was strong enough to induce some or all of the negative association or to mask a weaker positive (thus, harmful) effect, leading to the observation of a protective association. With only a signed DAG, it is not possible to tell which is correct, and the sensitivity analysis approaches described in later chapters would become essential. (text continues below)

If both arrows are positive (represented by + signs) or negative (represented by − signs), the likely direction of the net bias will be positive or upwards (represented by a + sign), because multiplying two numbers with the same sign will result in a positive number. If the two arrows have opposite signs, the likely direction of the net bias will be negative or downwards (i.e. towards the null for a positive association), represented by a − sign.

If the signs of the uncontrolled confounding and the observed (biased) study estimate are opposite, it could be concluded that the true bias-adjusted effect would have been in the same direction as observed in the biased study estimate. Such cases can still allow imperfect evidence to contribute informative information to support, rather than detract from, a given evaluation.

2.7 Use of DAGs in evidence synthesis

In the synthesis of the evidence across a number of studies with different study designs and different study populations, there is unlikely to be a single DAG that can describe the data-generation process in full. However, it can be helpful in evidence synthesis to begin with a working DAG that can be adapted to study-specific assessments to identify the potential limitations of each study and identify a set of variables that are likely to be necessary to control for confounding. It is also helpful for a group conducting evidence synthesis to work through the working DAG to ensure that assumptions are clearly understood between the group members and to identify areas of disagreement.

Table 2.6. Likely direction of confounding bias in the simplified scenario of a single uncontrolled confounder (*C*) for the directed acyclic graph (DAG) in Fig. 2.22, if the monotonicity assumptions for signed DAGs are met

Sign of arrow 1 from family history of colorectal cancer to red meat consumption ($C \rightarrow X$)	Sign of arrow 2 from family history of colorectal cancer to colorectal cancer ($C \rightarrow Y$)	Likely direction of confounding
+ (*C* increases risk of *X*)	+ (*C* increases risk of *Y*)	+ (positive[a])
− (*C* decreases risk of *X*)	− (*C* decreases risk of *Y*)	+ (positive[a])
+ (*C* increases risk of *X*)	− (*C* decreases risk of *Y*)	− (negative[b])
− (*C* decreases risk of *X*)	+ (*C* increases risk of *Y*)	− (negative[b])

C, uncontrolled confounder (family history of colorectal cancer); *X*, exposure (red meat consumption); *Y*, outcome (colorectal cancer).
[a] Positive uncontrolled confounding: not adjusting for *C* induces a positive association between *X* and *Y*, even when *X* does not affect *Y*.
[b] Negative uncontrolled confounding: not adjusting for *C* induces a negative association between *X* and *Y*, even when *X* does not affect *Y*.

For case–control studies, arrows from the outcome to the selection node will need to be included. For studies in which healthy worker biases are common, it may be essential to add nodes that describe the selection and confounding biases created as a result. Moreover, different measures used for different variables, or the timing of those measures, may lead to different information bias structures.

Researchers may find signed DAGs less useful for complex scenarios, for example in situations when they are trying to use signed DAGs and anticipate selection bias, non-monotonic effects, complex confounding structures, effect heterogeneity, and so on. Readers will find it helpful to refer to the more detailed discussions of signed DAGs in the literature (VanderWeele et al., 2008; Lipsky and Greenland, 2022). The following chapters provide information on other tools for understanding the direction of bias.

Finally, something that has not been mentioned yet is that DAGs can also be useful for non-traditional analyses of data from cohort or case–control studies, including the use of instrumental variable methods, Mendelian randomization approaches, and other quasi-experimental designs that may be used in triangulation processes for evidence synthesis (Swanson, 2015). In fact, some of the principles described in this chapter can help in reasoning about bias in those studies, too. For example, loss to follow-up can create a selection or collider stratification bias in such studies, and drawing a DAG can help to understand why (Swanson, 2019).

As noted previously, it is always challenging to draw DAGs that truly represent the underlying data-generation process. It may be helpful to consult a review of published DAGs for examples (Tennant et al., 2021). Because disagreements about the structure of the DAG can occur, it can be helpful to draw more than one DAG, to tease out the different assumptions that members of a group conducting evidence synthesis may have about a particular study. This can guide critical sensitivity analyses in evidence synthesis (Mathur and VanderWeele, 2020a, b, 2022).

2.8 Summary

DAGs make different assumptions about the data-generation process explicit, enable the identification of areas of disagreement in those assumptions between members of a group conducting evidence synthesis, and help to identify important sources of bias in the individual studies and the collective body of evidence being used to identify hazards. Working through DAGs collectively can create a motivation for additional bias analyses or sensitivity analyses that can be used to identify which sources of bias are most likely to matter in drawing conclusions about a particular hazard. DAGs also provide a systematic way of identifying critical variables for valid estimation of the effect of an exposure on an outcome. Thus, they provide a useful tool for hazard identification, as a place to communicate the working model used to make judgements about the quality of the underlying studies, and serve as a model for using the evidence presented in the most efficient way possible.

Key message

Table 2.7 presents the likely conclusions that can be drawn from the results of a study and the results of a simple signed DAG with a single confounder, about whether the exposure is likely to have an effect on the outcome.

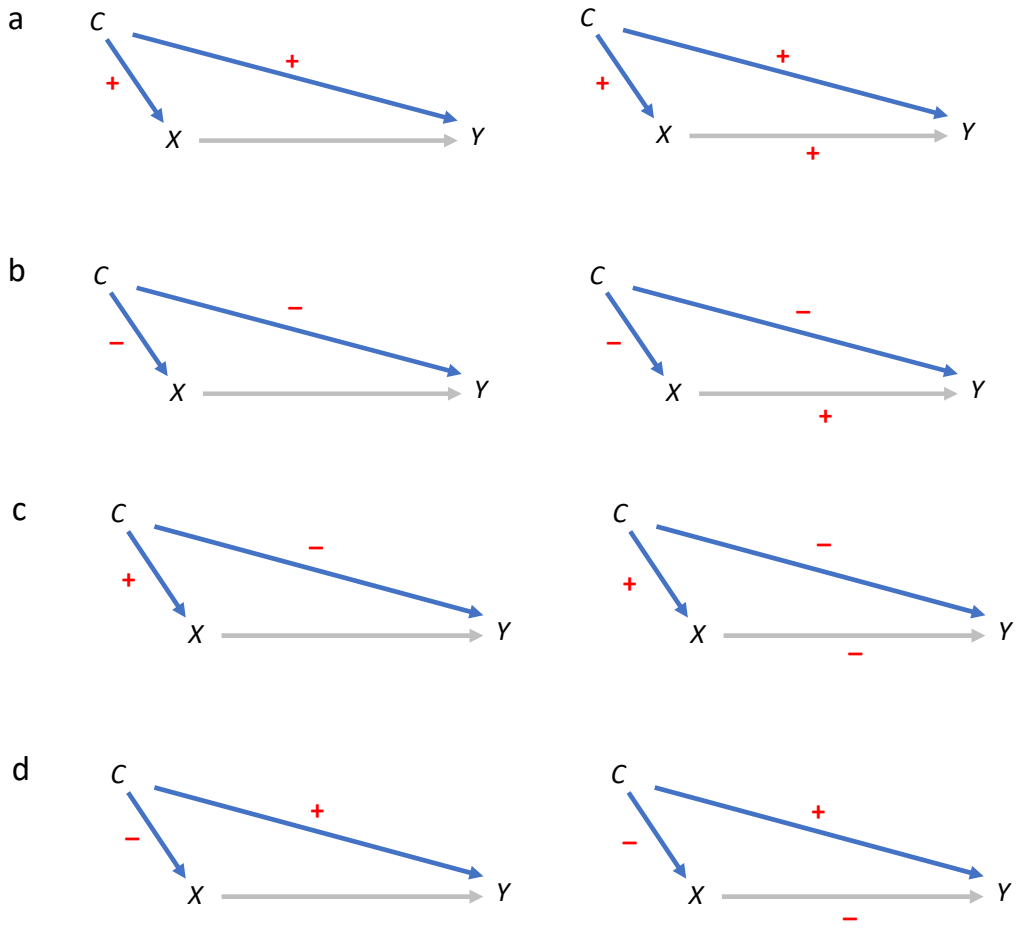
Table 2.7. Likely conclusions that can be drawn from a study about the existence of a non-null effect of an exposure on an outcome, based on the direction of confounding diagnosed with a signed directed acyclic graph (DAG) and the direction of the observed association

Observed association of exposure with cancer indicates	Signed DAG indicates that confounding is likely to be	Conclusion that can be drawn about the existence of a true (bias-adjusted) non-null effect (bias-adjusted RR, RD)
Elevated risk[a] (observed RR > 1, RD > 0)	Positive	*Unclear* Adjusting for the confounder would probably remove some or all of the estimate of the effect. Thus, it is not possible to say whether the estimate of the effect adjusted for the confounder would indicate increased, null, or decreased risk.
	Negative[a]	*Elevated cancer risk from exposure is likely* The observed estimate probably underestimates the true effect. Thus, adjusting for the unmeasured confounder would probably increase the estimate of the effect (bias-adjusted RR > observed RR; bias-adjusted RD > observed RD). Adjustment for the confounder would not bring the result back to the null or flip its direction.
No change in risk (observed RR = 1, RD = 0)	Positive	*Masked reduced cancer risk is likely* Adjusting for the unmeasured confounder would probably reveal a negative effect estimate (bias-adjusted RR < 1; RD < 0), indicating a probable reduced risk associated with the exposure.
	Negative	*Masked elevated cancer risk is likely* Adjusting for the unmeasured confounder would probably reveal a positive effect estimate (bias-adjusted RR > 1; RD > 0), indicating a probable elevated risk associated with the exposure.
Reduced risk (observed RR < 1, RD < 0)	Positive	*Reduced cancer risk from exposure is likely* Adjusting for the confounder would probably decrease the estimate of the effect. Uncontrolled confounding by this factor is unlikely to explain the observed result (i.e. adjustment for the confounder would not bring the result back to the null).
	Negative	*Unclear* Adjusting for the confounder would probably remove some or all of the estimate of the effect. Thus, it is not possible to say whether the estimate of the effect adjusted for the confounder would indicate decreased, null, or increased risk.

RD, risk difference; RR, relative risk.

[a] Indicates a scenario that would be most applicable to an *IARC Monographs* Working Group assessing whether an exposure could be carcinogenic (assuming positively coded exposure and cancer outcome variables, such that a positive exposure–outcome association with RR > 1 or RD > 0 would indicate harm).

Fig. 2.23. Possible results for the direction of bias as diagnosed with a signed DAG. The left side of each scenario shows the hypothesized direction (positive or negative) of the arrow, and the right side of each scenario depicts the likely direction (positive or negative) of the net bias in the *X*–*Y* relation.



It should be cautioned that DAGs that depict the full data-generation process, capturing information bias and selection bias, can make it seem impossible to approximate the causal effect. In some circumstances, this will indeed be true, but because the magnitude of the bias cannot be demonstrated in DAGs, it can be easy to think that all potential sources of bias are equal, are additive, and are severe, when in fact this may not be true. The following chapters discuss ways to identify the possible magnitude of the impact, so that sources of bias that have minimal impact can be ignored. Because DAGs do not represent the amount of bias created, they can lead to excessive concerns about some sources of bias. In such situations, bias analyses can help to sort out which sources are most likely to matter; thus, the DAG is only a first step.

References

Arah OA (2017). Bias analysis for uncontrolled confounding in the health sciences. Annu Rev Public Health. 38(1):23–38. doi:10.1146/annurev-publhealth-032315-021644 PMID:28125388

Arah OA, Chiba Y, Greenland S (2008). Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. Ann Epidemiol. 18(8):637–46. doi:10.1016/j.annepidem.2008.04.003 PMID:18652982

Berkson J (1946). Limitations of the application of fourfold table analysis to hospital data. Biometrics. 2(3):47–53. doi:10.2307/3002000 PMID:21001024

Brennan AT, Getz KD, Brooks DR, Fox MP (2021). An underappreciated misclassification mechanism: implications of nondifferential dependent misclassification of covariate and exposure. Ann Epidemiol. 58:104–23. doi:10.1016/j.annepidem.2021.02.007 PMID:33621629

Chavance M, Dellatolas G (1993). Bias from dependent misclassification. Epidemiology. 4(2):180–2. doi:10.1097/00001648-199303000-00015 PMID:8318094

Flanders WD, Waller LA, Zhang Q, Getahun D, Silverberg M, Goodman M (2022). Negative control exposures: causal effect identifiability and use in probabilistic-bias and Bayesian analyses with unmeasured confounders. Epidemiology. 33(6):832–9. doi:10.1097/EDE.0000000000001528 PMID:35895515

Fox MP, MacLehose RF, Lash TL (2021a). Misclassification. In: Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer; pp. 141–195. doi:10.1007/978-3-030-82673-4_6

Fox MP, MacLehose RF, Lash TL (2021b). Selection bias. In: Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer; pp. 75–103. doi:10.1007/978-3-030-82673-4_4

Glymour C, Scheines R (1986). Causal modeling with the TETRAD program. Synthese. 68(1):37–63. doi:10.1007/BF00413966

Greenland S (2005). Multiple-bias modelling for analysis of observational data. J R Stat Soc Ser A Stat Soc. 168(2):1–25. doi:10.1111/j.1467-985X.2004.00349.x

Greenland S, Pearl J, Robins JM (1999). Causal diagrams for epidemiologic research. Epidemiology. 10(1):37–48. doi:10.1097/00001648-199901000-00008 PMID:9888278

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

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

Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 155(2):176–84. doi:10.1093/aje/155.2.176 PMID:11790682

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300. doi:10.1177/003591576505800503 PMID:14283879

Kristensen P (1992). Bias from nondifferential but dependent misclassification of exposure and outcome. Epidemiology. 3(3):210–5. doi:10.1097/00001648-199205000-00005 PMID:1591319

Lash TL (2007). Heuristic thinking and inference from observational epidemiology. Epidemiology. 18(1):67–72. doi:10.1097/01.ede.0000249522.75868.16 PMID:17149141

Lash TL, Rothman KJ (2021). Selection bias and generalizability. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. Modern epidemiology. 4th ed. Philadelphia (PA), USA: Wolters Kluwer; pp. 315–332.

Lash TL, VanderWeele TJ, Rothman KJ (2021). Measurement and measurement error. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. Modern epidemiology. 4th ed. Philadelphia (PA), USA: Wolters Kluwer; pp. 287–314.

Lipsky AM, Greenland S (2022). Causal directed acyclic graphs. JAMA. 327(11):1083–4. doi:10.1001/jama.2022.1816 PMID:35226050

Mathur MB, VanderWeele TJ (2020a). Robust metrics and sensitivity analyses for meta-analyses of heterogeneous effects. Epidemiology. 31(3):356–8. doi:10.1097/EDE.0000000000001180 PMID:32141922

Mathur MB, VanderWeele TJ (2020b). Sensitivity analysis for unmeasured confounding in meta-analyses. J Am Stat Assoc. 115(529):163–72. doi:10.1080/01621459.2018.1529598 PMID:32981992

Mathur MB, VanderWeele TJ (2022). Methods to address confounding and other biases in meta-analyses: review and recommendations. Annu Rev Public Health. 43(1):19–35. doi:10.1146/annurev-publhealth-051920-114020 PMID:34535060

Pearl J (2009). Causality: models, reasoning, and inference. 2nd ed. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511803161

Pearl J, Mackenzie D (2018). The book of why: the new science of cause and effect. New York (NY), USA: Basic Books.

Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol. 88(5):879–903. doi:10.1037/0021-9010.88.5.879 PMID:14516251

Sheikh M, Shakeri R, Poustchi H, Pourshams A, Etemadi A, Islami F, et al. (2020). Opium use and subsequent incidence of cancer: results from the Golestan Cohort Study. Lancet Glob Health. 8(5):e649–60. doi:10.1016/S2214-109X(20)30059-0 PMID:32353313

Swanson SA (2015). Communicating causality. Eur J Epidemiol. 30(10):1073–5. doi:10.1007/s10654-015-0086-6 PMID:26445997

Swanson SA (2019). A practical guide to selection bias in instrumental variable analyses. Epidemiology. 30(3):345–9. doi:10.1097/EDE.0000000000000973 PMID:30896458

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

van Smeden M, Lash TL, Groenwold RHH (2020). Reflection on modern methods: five myths about measurement error in epidemiological research. Int J Epidemiol. 49(1):338–47. doi:10.1093/ije/dyz251 PMID:31821469

Vanderweele TJ, Arah OA (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 22(1):42–52. doi:10.1097/EDE.0b013e3181f74493 PMID:21052008

VanderWeele TJ, Hernán MA (2012). Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. Am J Epidemiol. 175(12):1303–10. doi:10.1093/aje/kwr458 PMID:22569106

VanderWeele TJ, Hernán MA, Robins JM (2008). Causal directed acyclic graphs and the direction of unmeasured confounding bias. Epidemiology. 19(5):720–8. doi:10.1097/EDE.0b013e3181810e29 PMID:18633331

VanderWeele TJ, Shpitser I (2013). On the definition of a confounder. Ann Stat. 41(1):196–220. doi:10.1214/12-AOS1058 PMID:25544784

Wright S (1960). The treatment of reciprocal interaction, with or without lag, in path analysis. Biometrics. 16(3):423. doi:10.2307/2527693

Yland JJ, Wesselink AK, Lash TL, Fox MP (2022). Misconceptions about misclassification: non-differential misclassification does not always bias results toward the null. Am J Epidemiol. doi:10.1093/aje/kwac035 PMID:35231925

Chapter 3. Confounding: a routine concern in the interpretation of epidemiological studies

3.1 Introduction 64
3.2 Evaluating control for confounding 65
3.3 Tools for assessing bias due to confounding 73
3.4 Summary 82

CHAPTER 3.

Confounding: a routine concern in the interpretation of epidemiological studies

David B. Richardson, Sadie Costello, Jay S. Kaufman, Kaitlin Kelly-Reif, Sarah Lewis, Kyle Steenland, and Eric J. Tchetgen Tchetgen

3.1 Introduction

As noted in the Preface, confounding arises when the exposure and the outcome of interest share a common cause. Informally, confounding may be described as a condition in which the association of exposure with the outcome is, in part, due to differences in outcome risk between the exposed and the unexposed that are not due to exposure effects on the outcome. A confounder is then defined as a variable that is responsible for confounding; typically, such a variable is a cause of the outcome that is associated with exposure but not affected by exposure. More precise definitions can be provided within formal causal models, such as potential-outcome and graphical models (Greenland et al., 1999a; Hernán and Robins, 2023; see also Chapter 2); these models will not be discussed here, but the reader is warned that there can be various definitions of confounding and confounders in these more formal discussions.

At an *IARC Monographs* meeting, the epidemiological studies under review are typically observational, meaning that the investigators did not have control over the exposure of interest (or any other variables) and, importantly for this chapter, did not randomly assign study participants to exposure. In observational studies, it is seldom reasonable to assume that pre-exposure factors that affect the outcome are equally distributed across subgroups defined by exposure; rather, exposure is often influenced by other factors, some of which may be risk factors for the cancer outcome of interest. Consequently, confounding is a common concern for Working Group members. Thus, one of the primary questions posed to reviewers in an *IARC Monographs* Working Group is "Can we reasonably rule out confounding as an explanation for an observed exposure–cancer association?"

A standard approach to the problem of confounding is to measure the important factors (e.g. pre-exposure factors that are predictive of the outcome in a cohort study) that may differ between exposure groups and to match on them in the study design (to the extent possible) or adjust for them in the analysis. If all the important confounders were accurately measured, an investigator might be able to obtain a valid estimate of the causal effect of the exposure on the outcome. However, the choice of which variables to control for (a judgement informed by causal, in

addition to statistical, considerations) is crucial because bias in an estimate of a defined exposure–disease association can be induced, or increased, by inappropriate control for covariates (Greenland et al., 1999b; Cole et al., 2010). Occasionally, IARC reviewers may encounter a study that used an approach intended to control for unmeasured as well as measured potential confounders. A classic example of such an approach is a randomized controlled trial, but other examples encountered in observational studies include analyses that leverage a natural experiment or an instrumental variable (such as genetic variation in a Mendelian randomization analysis; see Side Box 3.1). However, many epidemiological studies of cancer cannot or do not use these approaches; hence, uncontrolled confounding is often an important consideration for reviewers.

In Section 3.2, the reader will gain an understanding of how to evaluate control for confounding in published studies. Directed acyclic graphs (DAGs) (Chapter 2) will be referenced to represent assumptions regarding causal relations between variables and to assist in identifying causal effects. In Section 3.3, the reader will gain an understanding of approaches to assess potential bias due to uncontrolled confounding.

Given the focus on cancer studies, throughout the chapter confounding is considered as it applies to analyses of a binary outcome variable and ratio measures of association (such as rate ratios, odds ratios, hazard ratios, or risk ratios, as typical of most cancer studies). It is assumed that reviewers are interested in the total effect of the exposure on an outcome; therefore, mediation analysis, which is covered in VanderWeele (2016) and Hernán and Robins (2023), is not addressed here.

Chapter 2 introduced the use of DAGs to frame the identification and control of confounding. The focus in this chapter is on the evaluation of confounding within the context of a review that aims at hazard identification. Consequently, the focus is on whether uncontrolled confounding of a particular study result is a major source of bias and could meaningfully change a conclusion regarding that study's contribution for (or against) evidence of an association between the agent under review and the cancer outcome of interest. Evaluation of control for confounders is also commonly included in systematic reviews and meta-analyses through the use of tools to assess study quality. Such approaches rely on methods to assess the risk of bias due to confounding. While tools to assess study quality can be useful for helping a reviewer to think systematically about sources of bias, they are best used by substantive experts who can also consider the direction and magnitude of potential confounding and consider a range of methods to assess it. An uncritical use of risk-of-bias tools can lead to unwarranted dismissal of some studies because of alleged but unimportant confounding (Steenland et al., 2020). Methods are described in this chapter for an assessment of potential confounding bias, which may be useful when reviewing studies that inform an *IARC Monographs* evaluation. As described in Chapter 1, IARC has published general guidelines regarding the assessment of bias, and the methods outlined here are consistent with this guidance (IARC, 2019).

While this chapter focuses solely on confounding, there may be factors that are modifiers of the association under study (as well, perhaps, as confounders of it). In addition to considering whether a factor is a confounder, a reviewer might consider whether that factor modifies the association under study, meaning that the association on the selected measurement scale (e.g. relative risk) varies across values of the factor. Given a published report, a reviewer may be limited in such considerations by the information reported. For example, if the authors of a publication only report a covariate-adjusted estimate, then a reviewer cannot distinguish confounding by that factor from effect measure modification. However, if results have been stratified on a factor, and if the association varies importantly across strata of that factor, then there is modification of the association on that effect scale. Conversely, if the association is the same across strata, then the factor is probably not a modifier of the association (but could be a confounder of the association in a crude analysis that collapses information across strata of the factor). A variable can be a confounder, an effect measure modifier, both, or neither.

3.2 Evaluating control for confounding

When evaluating control for confounding in a published study (Fig. 3.1), reviewers will typically consider the following four topics (these will be explained in Sections 3.2.1–3.2.4):

- the study design;
- the study setting and restriction of the study population;

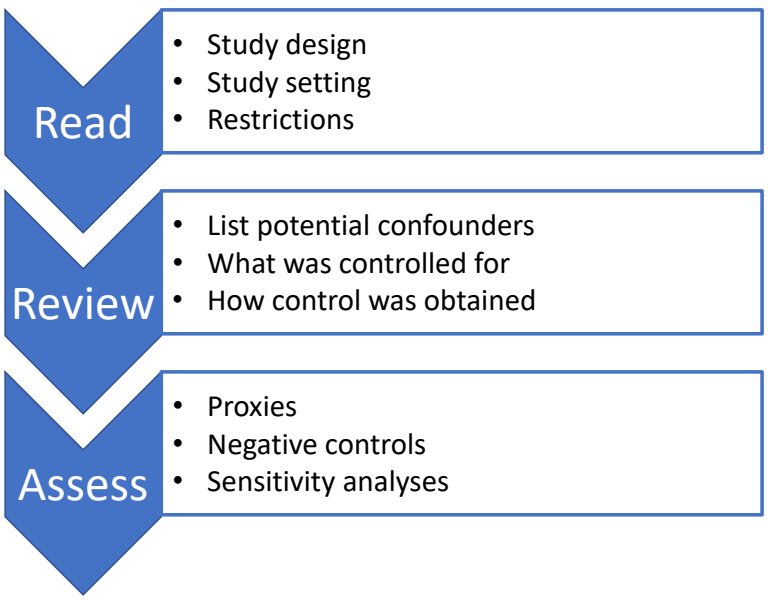
Side Box 3.1. Study designs or analyses and confounding considerations

Study design or analysis	Confounding consideration
Randomized controlled trial	Randomized trials are the most widely accepted method of addressing concern about confounding without necessarily measuring or adjusting for covariates. Nonetheless, randomized trials are rarely used in human subjects for evaluations of known or suspected carcinogens, because the administration of suspected carcinogens is unethical and because a long follow-up period is typically required in studies of cancer to observe the effect of exposure on cancer occurrence. Consequently, there are seldom many relevant randomized controlled trials to assess the carcinogenic potential of substances reviewed in *IARC Monographs*. Possible reasons for confounding in a randomized trial include imperfect allocation concealment, blinding, and adherence (i.e. compliance), as well as loss to follow-up; such considerations are important in studies of cancer outcomes because it is difficult to maintain adherence to a treatment protocol over many years.
Case-only (self-controlled)	Case-only designs, such as case-crossover and case-specular studies, are used to address concerns about potential confounding by characteristics that are constant over time. Only those participants who experience the outcome of interest are included, and study participants act as their own control. Because comparisons are made within individuals, confounding by characteristics that are constant over time is not possible. However, self-controlled designs do not typically lend themselves to investigations of cancer, which often feature long induction and latency periods. Possible reasons for bias in a self-controlled design include time-varying confounders and selection bias.
Matched fixed effects design (e.g. sibling or twin study)	Sibling and twin designs are used to address concerns about potential confounding by measured and unmeasured time-invariant factors. These studies involve pairs of participants who might be viewed as matched for a large number of potential confounders (e.g. genetics and childhood environment); these are shared or invariant characteristics within pairs and are often handled as fixed effects (intercepts) in a model. Possible reasons for bias in such designs include confounding by non-shared factors (i.e. those that may vary within a twin pair or sibling set) and selection bias (Frisell et al., 2012; Sjölander et al., 2022).
Instrumental variables, natural experiment, or Mendelian randomization analysis	Instrumental variables (IVs) are variables that are only associated with the outcome via their association with the exposure but are not affected by the exposure. A natural experiment is a type of IV analysis that involves settings in which the exposure variation for an individual (or a group of people) is due to an external factor, such as a natural disaster or an industrial accident, that is assumed to be related to the disease only through the exposure. A Mendelian randomization study is a type of IV analysis in which the investigators use genetic variation as the basis for a type of natural experiment. One important motivation for its use is that it offers the possibility to control for unmeasured confounders in an observational study (under certain strong identifying conditions). A Mendelian randomization analysis requires that the exposure under study has known genetic variants, which are strongly associated with it (an assumption termed *relevance*); it also requires that these genetic determinants are not associated with the outcome independently of the exposure (an assumption termed *exclusion restriction*), and it requires that the association between the instrument and the outcome is not confounded (an assumption termed *independence*) (Pierce et al., 2018). An increasing number of observational studies are using Mendelian randomization methods to study the effects of exposures, lifestyle factors, or biomarkers on cancer. If the necessary conditions do not hold, IV estimates may be biased (Hernán and Robins, 2006).
Cohort study	Cohort studies are commonly encountered in IARC evaluations of human carcinogens; these are observational in nature and are susceptible to confounding. This is particularly an issue when considering socioeconomic and lifestyle factors and cancer risk, because these factors tend to cluster and confounding within such studies can be substantial (Davey Smith et al., 2007). The collection of information on potential confounders in a retrospective cohort design may be limited by the available historical information, and that in a prospective cohort design may be limited by the knowledge at the time of study enrolment. Nested case–control studies (nested within cohorts) may be conducted for efficiency or to collect information on important confounders that may not have been collected in the original cohort study.

Side Box 3.1. Study designs or analyses and confounding considerations (continued)

Study design or analysis	Confounding consideration
Case–control study	Case–control studies often involve the collection of detailed information on at least a few potential confounders that are of primary focus; they typically focus on one outcome and are usually smaller than cohort studies, allowing for a richer collection of data on risk factors for the single outcome of interest and other covariates than a cohort study. Matching on potential confounders, more commonly observed in case–control studies than in cohort studies, can improve efficiency when adjusting for confounders (and may permit control for confounding that is otherwise difficult to achieve, as in neighbourhood matching). Within case–control studies, there may be issues with the control participants not being representative of the population from which the case participants arose, which could introduce confounding; also, confounders may not be measured well if individuals are asked to recall lifestyle factors that occurred before cancer symptoms were observed. In addition, confounders may be recalled differently by case and control participants (selection bias and information bias; see Chapters 4 and 5). From a confounding perspective, nested case–control studies have similar issues to cohort studies.
Ecological study	These are studies in which the exposure is studied at a population level rather than an individual level, and variation in outcome is examined in relation to variation in population prevalence of exposure. Ecological studies often have limited or no information on individual-level confounders and are consequently susceptible to confounding (including a particular form of bias that may arise in ecological study analyses because of confounding or effect measure modification between groups under comparison).
Cross-sectional study	Such studies typically play a minor role in cancer evaluations. Confounding is an issue; cross-sectional studies often have the additional complexity of temporal ambiguity. It may be unclear whether the exposure preceded the disease; it may also be unclear whether a covariate preceded the exposure and thus whether it is a confounder. (text continues on page 65)

Fig. 3.1. Steps to take when assessing confounding in individual studies.



- the set of covariates that were adjusted for in the analysis (and how those covariates were measured and modelled); and
- important confounders that were not controlled for.

3.2.1 Study design

The study design is an important starting point for evaluating control for confounding; it is possible to control for confounding in the study design phase. The choice of study design may direct a reviewer's attention to certain key areas for consideration, such as the appropriateness of an external comparison group for the analysis of standardized mortality ratios in an occupational cohort study. It may even obviate the need to focus attention on the adequacy of control for certain types of confounders. For example, matched designs involving siblings born of the same mother are sometimes used to control for maternal factors that remain constant between pregnancies, such as maternal genetics and some aspects of lifestyle and socioeconomic status (see Side Box 3.2).

3.2.2 Study setting and restrictions

A careful decision regarding study setting can help to minimize confounding, for example by finding populations that lack an association between a confounder and the exposure of concern. For instance, a large cohort of Seventh-Day Adventists offers a setting with little or no confounding by alcohol consumption or smoking, because these behaviours are largely absent in that population (Butler et al., 2008). Similarly, restriction of the study population (e.g. by sex, geography) can help to control for confounding. Sometimes restriction on a confounder can provide control over factors that would otherwise be difficult to measure and control for in an analysis. For example, restriction to a single continental population, such as Europeans (Auton et al., 2015), to minimize population stratification (confounding by ancestry) is common in genome-wide association studies (although many contemporary genome-wide analyses also adjust for finer population structure). As another example, occupational cohort studies are often conducted in a setting in which the workers involved share similarities in terms of education, income, access to medical care, geography, and lifestyle factors (e.g. diet). Consequently, in occupational studies with internal comparisons, such factors are usually of less concern as confounders than they are in environmental studies, because these lifestyle factors should have limited associations with occupational exposure.

However, inappropriate restriction can lead to bias (e.g. if restriction is on an intermediate or mediating variable or collider; see Chapters 2 and 5). Moreover, restriction necessarily affects the generalizability of results (and reduces sample size), so it should be carefully assessed.

3.2.3 Covariates that were (and were not) adjusted for in a published analysis

A standard approach to addressing the problem of confounding is to measure important factors that may differ between exposure groups and adjust for them in the analysis. Here, the focus is on analyses where the aim is to control for confounding by adjustment for measured variables (e.g. adjusting for the variable in a regression model for the outcome).

It is important to consider both the confounder–outcome association and the confounder–exposure association. Those involved in an expert review, such as an *IARC Monographs* evaluation, will often come to a consensus on the important potential confounders of an association under evaluation. One source of information about such potential confounders is the study publications under review; authors often provide useful guidance in their publications about measured and unmeasured potential confounders, as well as omitted potential confounders. However, regardless of the authors' description of important potential confounders, reviewers may have a different view. Authors often describe their approach to the final selection of their covariate adjustment set; again, regardless of how the adjustment variables were selected in a given publication, the reviewers' responsibility at this stage is to assess whether the important potential confounders have been sufficiently controlled for.

Reviewers may wish to start by considering the confounder–outcome association, focusing on those factors that are established causes of the cancer outcome under study. Useful sources of such information are the *IARC Monographs* and the IARC list of classifications of agents for which there is *sufficient* and *limited* evidence of carcinogenicity in humans by cancer site; similarly, the *IARC Handbooks of Cancer Prevention* can provide information on potential

Side Box 3.2. Some approaches to control for confounding

Matching in the design of a study can sometimes allow for control for factors that would otherwise be difficult to adjust for efficiently in the analysis (e.g. in the absence of a matched design, because of sparse data).

Matching may be used in cohort or case–control studies. In a matched cohort study, an investigator might enumerate an unexposed group of study participants who match the exposed study participants in terms of some characteristics (such as age and sex) that are of concern as potential confounders; a comparison of the occurrence of cancer between the exposed and unexposed groups will not be confounded by those factors that were matched on in the design. Matching is often used in case–control studies of cancer outcomes, with the aim of improving efficiency in a case–control analysis when it would otherwise be necessary to adjust for a matching factor, such as attained age. Similarly, in a population-based case–control study of a rare cancer, neighbourhood matching of case and control participants may allow for adjustment for characteristics that are shared by neighbours, such as socioeconomic, diet, or lifestyle factors, but that may be difficult to adjust for in the analysis in the absence of such matching, because of sparse data or difficulty in obtaining sufficient or accurate data to control for such hard-to-quantify variables. In certain settings, self-matching can be used (e.g. the case–control status is determined by the location of the tumour in relation to the exposure within the body, as in Example 3.1). However, as noted in Chapter 2, an important difference from matching in cohort studies is that case–control matching is a form of selection bias that distorts associations and trends (Mansournia et al., 2018). To control this bias, the analysis must include adjustment for the matching variables in a form at least as detailed as the form used for matching; this means, for example, that if age matching is done in 5-year categories, then the adjustment must use age as a categorical variable with categories at least as narrow as 5 years.

Example 3.1. Self-matching to control for confounding

A case-only study (Maclure, 1998) was conducted of mobile phone use and glioma (Larjavaara et al., 2011). The location of the actual tumour site (i.e. the case site) was compared with a control site, defined as the mirror image site obtained across the midpoint of the axial and coronal planes of the patient's brain (i.e. within the same person). The control sites were effectively matched to the case sites on each pair being within the same patient's brain. The case and control sites were then compared with respect to estimated mobile phone exposure, to determine whether the phone was used on the side of the brain where the tumour occurred. In this design, participants with cancer each served as their own control; therefore, confounding by personal characteristics (such as age, sex, income, or diet) was judged to be unlikely in these analyses.

Side Box 3.2. Some approaches to control for confounding (continued)

Other approaches to study design that are sometimes used in cancer research to address potential confounding involve leveraging situations in which exposure was determined by factors beyond the control of the investigator but that arguably mimic random exposure assignment. Such studies are sometimes called natural experiments or quasi-experimental designs, as explained in Example 3.2.

In a natural experiment, the assignment mechanism is a form of instrumental variable (IV), because it influences exposure but only influences the outcome through its effect on the exposure. Quasi-experimental designs have been used in evaluations of interventions on tobacco, air pollutants, and petrochemical exposures. One version of IV analysis that is sometimes encountered in epidemiological studies of cancer outcomes is Mendelian randomization, in which the IV is the random inheritance of genetic variants that are known to predict exposure, under the classic assumption that genetic factors are inherited independently of each other (note that this assumption may not hold for genetic variants that are located near one another on the same chromosome). Genetic variants are usually not subject to confounding by lifestyle and environmental factors (Smith and Ebrahim, 2003). In Mendelian randomization analyses, populations are grouped according to the presence of genetic variants (alleles) that are associated with the exposure of interest. Comparison of cancer risk between genetic groups that are associated with the exposure can provide an unconfounded estimate of the effect of the exposure on cancer (Yarmolinsky et al., 2018). (text continues on page 68)

Example 3.2. An example of a natural experiment to control for confounding: a military conscription lottery

A situation that has been used in a natural experiment is a military conscription lottery (where one compares those drafted with those not drafted for cancer outcomes, such as in studies that have examined effects of service in the Viet Nam era on cancer occurrence).

confounders that are cancer-preventive factors (see Section 6.3.1 for more examples). For a study of a given cancer outcome, a reviewer can readily refer to such lists of known or suspected causes of that cancer to inform consideration of potential confounders. Note that because confounder–outcome associations are rarely homogeneous from one cancer site to another, the list of potential confounders of concern will also vary by cancer site. As noted in Chapter 1, the evaluation of human evidence regarding carcinogenicity is also specific to each cancer site; therefore, concern about a potential confounding factor (e.g. smoking) might be reasonably excluded for certain cancer types (e.g. melanoma) but not others (e.g. lung cancer).

In addition to the confounder–outcome association, it is also necessary to consider the confounder–exposure association. An important consideration is whether potential confounders precede the exposure of interest. Therefore, reviewers may often rely on information on the distribution and determinants of exposure. A reviewer may encounter situations in which adjustment was made for a covariate that was measured after the exposure of interest occurred. In such situations, careful consideration should be given to whether exposure influenced that covariate; however, there are settings for which an investigator may reasonably assume that the measured value of such a covariate is a good approximation of its pre-exposure value and is unaffected by the exposure of interest (e.g. educational attainment, assessed after exposure, in a study of the effect of an exposure in a population of middle-aged adult patients). The factors that influence exposure to an agent may vary over time and between populations and may depend on economic and social factors, laws and regulations, and social and behavioural factors. Consequently, in assessment of confounding, information should be obtained and used on how the association of a potential confounding factor with the exposure and the disease may vary across different study populations.

A reviewer of a published article will consider whether any important potential confounders were not accounted for (e.g. not controlling for smoking in a study of a given exposure in relation to lung cancer). However, the fact that a variable that a reviewer posited as a potential confounder was not adjusted for in a published study does not necessarily mean that it was a strong confounder (or even a confounder at all). Often the authors of a publication will describe the rationale for exclusion of a variable from the adjustment set and may report results that were obtained with different sets of adjustments for covariates. As shown in Example 3.3, a factor could be an established cause of cancer but might not confound the association of interest in the population under study.

Another consideration in a review of a published article is whether any of the variables adjusted for in the published analysis were not potential confounders but rather could induce or exacerbate confounding through inappropriate control. The term *overadjustment* is sometimes used to refer to bias induced by adjustment for intermediate variables or variables downstream from exposure – to use the language described in Chapter 2, to disrupt a chain from exposure to outcome. Adjusting for a variable that is on the causal pathway is an example of overadjustment (Schisterman et al., 2009). Overadjustment can also sometimes refer to a different problem: the bias (or loss of precision) that can occur in an analysis that controls for a strong predictor of exposure that is not associated with the outcome. In some settings, adjustment for a strong predictor of the outcome that is not associated with exposure also can induce a form of overadjustment bias, because such adjustment may push an estimate of the log odds ratio away from the null (Greenland et al., 2016). To help fully understand and discuss potential confounders, a diagram, such as a DAG, showing presumed causal relations among variables (and their measurements) can represent the assumed underlying causal associations and any confounding pathways implied (see Chapter 2), as shown in Example 3.4.

3.2.4 How confounders were measured and modelled

Consideration of how the confounders included in an adjustment set were measured and modelled is important because it relates to concerns about residual confounding by the factor after adjustment. Imperfect measurement of a confounding variable will usually lead to incomplete control of confounding (i.e. residual confounding) that is proportional to the amount of confounding originally present (Greenland, 1980; Greenland and Robins, 1985; Savitz and Barón, 1989; Ogburn and VanderWeele, 2012). For example, smoking may be imperfectly controlled in an analysis that classifies whether a person has ever smoked but does not account for whether the person is a current smoker or a former smoker or for the amount and duration of smoking. If the amount of original confounding was substantial, then – regardless of the fraction that was controlled – the amount that was not controlled may

Example 3.3. Adjustment for body mass index in studies on red meat consumption and colorectal cancer

In a meta-analysis (IARC, 2018), it was noted that many studies did not adjust for body mass index (BMI) because estimates of the association between red meat consumption and colorectal cancer (CRC) did not change after adjustment for BMI, although it is considered a potential confounder in the literature (Chan et al., 2011). Some may consider BMI to be a mediator on the pathway between red meat consumption and CRC (e.g. Example 2.1a), but in much of the literature BMI is considered to be a confounder that can affect both red meat consumption (those with higher BMI are likely to eat more red meat) and risk of colorectal cancer. For the Working Group's deliberations regarding the association between red meat consumption and colon cancer, the observation that inclusion of BMI in a regression model does not change the estimate of the association between red meat consumption and colon cancer suggests that BMI is neither an important mediator nor a confounder. (text continues above)

Example 3.4. Overadjustment as a concern in studies on shift work and cancer

In the *IARC Monographs* evaluation of the literature on night shift work in relation to breast cancer, the reviewers considered confounding and adjustment for other covariates (IARC, 2020). These considerations were particularly important because day workers are usually taken as the reference group and there may be many important differences between day workers and night workers with respect to risk factors for breast cancer. The Working Group consulted the literature to determine the degree to which lifestyle factors of day workers and night workers differ, to help in deciding whether a particular covariate was a potential confounder. For example, the reviewers noted that reproductive factors (parity, age at first birth, and menopause) are considered risk factors for breast cancer. They then cited studies that found differences between day workers and night workers with regard to reproductive factors; however, they noted that these associations were not strong. One could conclude that reproductive factors are potential confounders for the association between night shift work and breast cancer, but they are not likely to be strong confounders. Thus, a study that did not include reproductive factors may not suffer from much confounding bias. The reviewers also cited references indicating that several risk factors for breast cancer may be affected by night shift work, including disrupted sleep, physical activity, eating behaviours, and consumption of alcohol. Note that the total effect of night shift work on breast cancer includes the effect mediated by other factors. For example, perhaps night shift work increases the risk of breast cancer because people who work at night experience work-induced changes to exercise and diet that, in turn, lead to breast cancer. This does not imply that one would need to adjust for diet or exercise to obtain an unbiased result. In fact, the opposite is true: to obtain an unbiased result for the total effect of night work on breast cancer, one should not disrupt the causal chain by adjusting for behavioural factors that are affected by night work. (text continues on page 71)

still be important in absolute terms. Conversely, if a covariate is a weak confounder, residual confounding will have only a minor influence on the estimate of association. For example, smoking might be imperfectly controlled through next-of-kin reporting about whether a patient with breast cancer had ever smoked, but the residual confounding might be minor, given the weak smoking–breast cancer associations. Theoretically, in the extreme case of a very poorly measured confounder that suffers from systematic misclassification, adjustment for such an error-prone variable can make confounding worse (Ogburn and VanderWeele, 2012); however, such a scenario is typically implausible, and in most applications adjustment for an error-prone measure of a confounder will not make confounding worse (Greenland, 2012).

Importantly, some types of confounding are more difficult to control for than others. For example, specific exposures, such as tobacco smoking, lend themselves to careful measurement, whereas other factors that might confound an association of interest, such as socioeconomic conditions or health behavioural factors that influence exposure and cancer detection, are often almost impossible to measure well and fully control for in an analysis. If there are major differences at the outset (e.g. in a between-country comparison of breast cancer incidence rates), an investigator may have adjusted for a large set of covariates, yet the reviewers may remain sceptical that important confounding factors were adequately controlled. Another example of confounding that may be difficult to control arises in occupational studies when co-exposure occurs in the workplace to multiple correlated agents that could be carcinogenic (see Example 3.5).

(a) Time-varying confounders and time-varying confounders affected by prior exposure

So far, the discussion has been limited to confounding at one point in time, implying that study authors are interested in estimating the effect of an exposure that occurred at one time point on cancer. Many cohort studies involve the analysis of data

Example 3.5. Examining confounding by co-exposures in the workplace

In the *IARC Monographs* evaluation of the literature on night shift work in relation to breast cancer, the reviewers considered the association between occupational circadian rhythm disruption and breast cancer incidence among female flight attendants (IARC, 2020). Metrics of circadian rhythm disruption included employment duration, hours flying in the standard sleep interval, and number of time zones crossed. A potential confounder of concern was occupational exposure to cosmic radiation, which was highly correlated with employment duration. In the context of the IARC review, the concern was primarily with respect to positive confounding of the association between circadian rhythm disruption and breast cancer incidence; given the lack of observed association between circadian rhythm disruption and breast cancer incidence among female flight attendants (Pinkerton et al., 2016), the concern was not substantiated. However, in many situations when the primary exposure of interest is highly correlated with a potential confounder, reviewers may express concern about the ability to estimate the effect of the exposure of interest with adequate control for the confounding factor. (text continues on page 72)

on exposure at multiple different time points, and authors are frequently interested in the effect of lifetime or cumulative exposure (or possibly a lagged metric of cumulative exposure) on cancer. When exposures vary over time, so can confounding in an analysis that allows for time-dependent exposures. In the most straightforward scenario, a predictor of the outcome might also predict exposure at each time point. For example, in studies of occupational exposure on cancer, age could be a time-dependent confounder. In this situation, time-varying confounders can usually be treated in a standard manner; for example, in a regression model, a term for the confounder might be included at each time point.

However, some time-varying confounders can be affected by prior exposure. Reviewers of papers can evaluate the plausibility of confounders at a given point in time being influenced by prior exposure. Consideration of whether time-varying confounders are affected by prior exposure is important because in such situations standard outcome modelling of the associations is not guaranteed to yield unbiased results, as shown in Example 3.6 (Cook et al., 2002; Hernán and Robins, 2023).

3.3 Tools for assessing bias due to confounding

Control for confounding is rarely, if ever, sufficient to remove bias entirely. Rather, control is a matter of degree and often warrants a critical assessment of whether the control achieved in a published paper may be adequate to make a reasonable judgement regarding the effect of the exposure on cancer. After having reviewed the control for confounding in a study, a reviewer might suspect that the published analysis suffers from substantial confounding by uncontrolled covariates or suffers from residual confounding due to inadequately controlled covariates (e.g. confounders that were poorly measured, inadequately modelled, or poorly specified).

Given concern about possible confounding of an exposure–cancer association that was reported in an individual study, the next step is to assess the direction and magnitude of the confounding bias. This can help to understand the impact of uncontrolled confounding on the evidence under review.

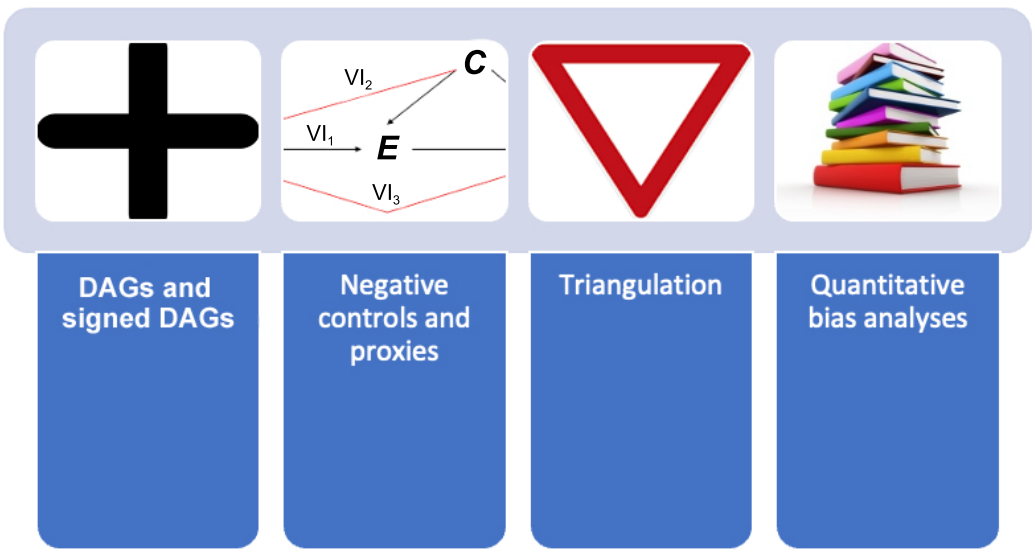
Various approaches (tools for assessing confounding, numbered C-# below) are available to inform evaluations of confounding by unmeasured variables of an observed association between cancer and exposure to an agent under evaluation in an IARC review. Investigators should consider the following, which will be developed further in subsequent subsections (Fig. 3.2):

- Tool C-1: DAGs and signed DAGs (i.e. causal relations between variables based on substantive knowledge);
- Tool C-2: negative control outcomes (or exposures) and proxies (evidence of confounding within a study);
- Tool C-3: triangulation (evidence of confounding between studies that differ meaningfully); and

Example 3.6. Healthy worker survivor bias

In occupational studies of carcinogens, a common concern is a form of confounding that is often referred to as healthy worker survivor bias, whereby workers who are less susceptible to the health effects of the exposure survive longer in the workplace and therefore may accrue more cumulative exposure. In this instance, leaving work is a time-varying confounder that is affected by prior exposure because (i) leaving work may predict cancer diagnosis (e.g. if a person left work for cancer-related reasons), (ii) leaving work will affect accrual of occupational exposure, and (iii) prior exposure to hazardous material may affect a worker's current employment status. Such a bias may occur if people who work nights quit when they cannot tolerate the lifestyle anymore. If years of night work have already taken a health toll on the worker in ways that are on the pathway to cancer, there may be time-varying confounding affected by prior exposure. Including time-varying explanatory variables in an outcome regression model for factors such as whether individuals are currently employed or the duration of each person's employment will not remove the bias. Special methods, known collectively as g-methods (generalized methods), are needed to address this issue and produce unbiased estimates (Robins, 1986; Hernán and Robins, 2006, 2023; Buckley et al., 2015). These g-methods enable researchers to model long-term exposure in a different way from ordinary outcome regression modelling. For example, rather than estimating risk from cumulative exposure over many years, g-computation, one type of g-method, estimates the risk of cancer from exposure in each year and then sums up the risks of cancer over time. This approach allows the researcher to adjust only for confounders that precede exposure in each year, thus addressing the bias from the healthy worker survivor effect. Unless g-methods are used, estimates of occupational exposure–cancer associations from studies that are affected by healthy worker survivor bias will typically be attenuated (i.e. biased downwards). (text continues on page 73)

Fig. 3.2. Tools to consider when evaluating the impact of probable confounding. Each of the approaches proposed requires substantive expertise, which may include expert judgement, information derived from internal substudies, or findings from external studies.



- Tool C-4: quantitative bias analyses.

For simplicity, the focus in this chapter is on assessment of the impact of a single primary confounder of concern. Often, for clarity in a review, it is useful to focus on assessment of the impact of one key potential confounder of concern at a time. Section 6.4 discusses some approaches for multiple-bias analysis, where more than one confounder (or other source of bias) is of concern.

Most of the approaches described here are premised on the ability to explicitly name a factor of concern as a confounder. This requires hypothesizing why that factor is associated with both the exposure and the outcome. Like substantive hypotheses, hypotheses about why a confounding factor is associated with exposure, and with disease, should be specific, should describe substantively important associations, and should make quantitative predictions of the confounding effect (Hertz-Picciotto, 2000). Given well-specified hypotheses about confounding of observed associations, a reviewer may be able to assess the degree to which results from observed data are likely to be substantially affected by the hypothesized confounding. A review is strengthened by explaining which factors were considered as potential confounders and why, as well as their likely effects (see Example 3.7).

For known confounders that have not been measured, it may be feasible to perform a bias analysis to suggest the possible effect of the unmeasured confounder. Of course, it might be the case that a reviewer does not wish to posit (or name) a specific confounder but rather wishes only to express a general concern that an observed association between exposure and disease might be confounded by a factor as yet unknown (at least to the investigator). A general concern about uncontrolled confounding might arise if a reviewer were to conclude that the important risk factors for a given cancer outcome have simply not yet been identified. In general, vague statements regarding entirely unknown confounders are less amenable to evaluation using most of the approaches described here. The less that is understood about disease etiology or exposure assignment, the greater the potential for unknown factors to be important confounders. The latter threat to validity can be minimized by focusing a hazard identification on a well-defined exposure (e.g. benzene) rather than a vague exposure or contextual factors (e.g. green space) (Hernán, 2016). Again, a Working Group's discussion of the role of confounding when evaluating evidence regarding the carcinogenicity of an agent will be most informative when the confounding factor is explicitly named, and when hypotheses regarding why that factor is associated with exposure and disease can be discussed and evaluated. Quantitative bias analysis can be used to assess whether the study results are sufficiently robust to render uncontrolled confounding unlikely (see Section 3.3.4).

3.3.1 Tool C-1: DAGs

As noted in Chapter 2, a simple DAG can serve as a starting point for the analysis of uncontrolled confounding. The drawing of a DAG requires substantive knowledge about covariates and their causal relations to the exposure and outcome of interest. Without such substantive knowledge, a DAG is largely speculative. Although a DAG is not an oracle that can provide infallible identification of confounding in a particular study, given substantive expertise (which often exists in expert Working Groups), it can be useful for reasoning about systematic bias and making the causal assumptions of Working Group members involved in an *IARC Monographs* evaluation explicit and clear.

A signed DAG (i.e. one in which the direction of the effect of a confounder is specified) can aid Working Group members in assessing the probable direction of bias due to confounding (see Section 2.6). Also, DAGs can inform the assessment of time-varying confounders. For

Example 3.7. Relative importance of confounders

In its examination of the carcinogenicity of red meat (IARC, 2018), the Working Group specified which confounders were thought to be important (physical activity, BMI, caloric intake) and gave more weight to studies that controlled for these confounders (or that demonstrated that adjustment for the covariate of concern did not have meaningful impact on the estimate of interest). (text continues above)

example, in an occupational cohort mortality study to investigate a suspected carcinogen, healthy worker survivor bias is a common concern. A signed DAG can help to judge whether this form of confounding is likely to be present; relevant considerations include the need for associations between (i) prior exposure and employment status and (ii) employment status and mortality (Naimi et al., 2013). A DAG can help to answer these questions and guide a reviewer's assessment of the likelihood of such bias.

3.3.2 Tool C-2: negative control outcomes (or exposures) and proxies

Sometimes a reviewer is able to indirectly assess confounding by an unmeasured factor using evidence available from within the published study, based on approaches that involve negative controls and proxies. These methods all share similar assumed causal structures between variables (Fig. 3.3). However, as discussed next (Sections 3.3.2(a) to 3.3.2(c)), Working Group members may find useful conceptual distinctions between negative control outcomes, negative control exposures, and proxies for an unmeasured confounder.

(a) Negative control outcomes

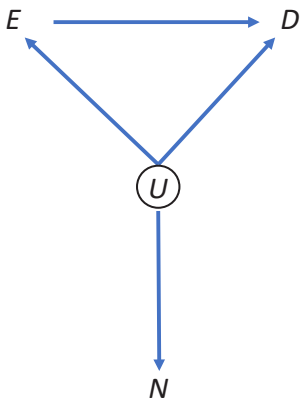
Suppose that a reviewer is concerned about potential confounding in a cohort study of the association between a suspected carcinogen and a site-specific cancer, but that the potential confounder was unmeasured in the study under review. A negative control outcome approach proceeds by examining the association between the suspected carcinogen and another outcome that (i) is caused by the hypothesized confounding factor and (ii) is not caused by the suspected carcinogen of interest.

Fig. 3.3 illustrates the causal associations described: *E* denotes the exposure of interest, *D* the outcome, *U* the unmeasured confounder, and *N* the negative control outcome. Note that *U* has a causal effect on *N* but *E* does not.

Under these conditions, an observed association between *E* and *N* would be entirely due to confounding by *U*. Therefore, the absence of an association between *E* and *N* would argue against the hypothesis that the *E*–*D* association is confounded by *U* (Example 3.8).

This approach is well suited to the evaluation of cohort studies where information on many outcomes (e.g. cause-specific mortality) has been collected; this may enable an investigator to examine not only the association between the exposure of interest and the outcome of primary interest but also the association between that exposure and an outcome that a reviewer posits as a useful negative control outcome. Absence of evidence of an association between *E* and *N* would help to nullify claims of confounding by *U*. This can be thought of as an example of internal (i.e. within-study) triangulation of evidence, where the examination of associations between exposure and outcomes with different presumed causal structures can be compared to indirectly assess bias (Pearce et al., 1986).

Fig. 3.3. Diagram for analyses involving a negative control or proxy, *N*. *E*, exposure; *D*, outcome; *U*, confounder.



Example 3.8. Negative control outcomes

In an investigation of the effect of red meat consumption (*E*) on cancer (*D*), where tobacco smoking (*U*) is not measured but is considered a potential confounder, an investigator might posit emphysema as a valid negative control outcome (*N*). If that assumption were correct, the absence of an association between red meat consumption and emphysema would be evidence that tobacco is not a confounder of a red meat consumption–cancer association. (text continues above)

A related approach is sometimes used in the interpretation of standardized mortality ratio (SMR) analyses to, as it were, correct SMRs for bias. The SMR for an outcome that is presumed to be susceptible to the same confounding factors as the outcome of primary interest, but is presumed not to be strongly associated with the exposures of interest, serves as a measure of the bias due to confounding. This approach has been used both qualitatively, to indirectly assess confounding when interpreting cause-specific SMRs, and quantitatively, to derive an adjusted SMR for the outcome of interest (and associated confidence interval) by taking a ratio of the measures.

As shown in Example 3.9, expert groups can quantitatively evaluate uncontrolled confounding by calculating an adjusted SMR using published results if appropriate negative control outcomes can be identified (Side Box 3.3) and are reported.

(b) Negative control exposures

Suppose that a reviewer is concerned about potential confounding of the association between a suspected carcinogen and a site-specific cancer, but that the potential confounder was unmeasured in the study under review. A negative control exposure approach proceeds by examining the association of the site-specific cancer outcome of interest with another exposure variable that (i) is associated with the hypothesized confounding factor and (ii) is not a cause of the site-specific cancer outcome of interest.

Fig. 3.3 can also illustrate the causal associations required for a valid negative control exposure if *N* is now taken to denote the negative control exposure: *N* shares common cause *U* with *E*, but *N* does not cause *D*.

Under these conditions, an observed association between *N* and *D*, adjusted for *E* (or within a stratum of *E*), would be entirely due to confounding by *U*, whereas the absence of such an association between *N* and *D* would be evidence against confounding of the *E*–*D* association by *U* (Example 3.10 and Side Box 3.4).

(c) Proxies for a confounder

Proxies are indirect measures of unavailable variables of interest; this chapter focuses on proxies that are used as surrogates for potential

Example 3.9. Indirect adjustment of SMRs to reduce healthy worker biases in aluminium smelting work

In a study of bladder cancer among workers in an aluminium smelting plant, confounding through healthy worker biases was a concern (McClure et al., 2020). The investigators quantitatively evaluated healthy worker effects through negative control outcomes and derived an adjusted SMR. They did this by selecting a group of diseases (e.g. non-malignant blood disorders, diabetes, psychological disorders) that satisfied the conditions of a negative control outcome because they were thought to be unaffected by smelting work exposure but would be affected by healthy worker effects in a fashion similar to bladder cancer. The unadjusted SMR for bladder cancer was 2.27, and the unadjusted SMR for the negative control group was 0.65. The adjusted SMR, derived by taking the ratio of the two SMRs, was 3.47; this indicated that the confounding from healthy worker effects downwardly biased the SMR for bladder cancer. (text continues above)

Side Box 3.3. Information needed to facilitate use of negative control outcomes to evaluate confounding

Several elements are required to use negative control outcomes to evaluate confounding. The first requirement is for a suitable negative control outcome, i.e. an outcome that is related to the confounder but is not caused by exposure to the agent under evaluation. Notably, the negative control outcome may be identified by the expert reviewer but not by the original researchers. Required results include the association between the agent of interest and the negative control outcome, as well as the primary association between the agent and the outcome of interest. (text continues above)

Example 3.10. Negative control exposures

In studies that assess exposure information by questionnaire, investigators will often include questions about exposure to agents that are thought to be unrelated to the outcome of interest; these may serve as negative control exposures. (text continues on page 77)

Side Box 3.4. Information needed to facilitate use of negative control exposures to evaluate confounding

Several elements are required to use negative control exposures to evaluate confounding. The first requirement is for a suitable negative control exposure, i.e. an exposure that is related to the confounder but is not a cause of the disease outcome under evaluation. As with the negative control outcome, the negative control exposure may be identified by only the expert reviewers. Required results include the negative control exposure–disease association, adjusted for exposure 1 (or negative control exposure–disease association within a stratum of exposure 1) between the negative control exposure and the outcome of interest, adjusted for the exposure of interest (or the negative control exposure association with the disease of interest within a stratum of the main exposure of interest), as well as the primary association between the agent and the outcome of interest. (text continues on page 77)

confounders. Here, a proxy is taken to be a variable associated with an uncontrolled confounder *U* that would be irrelevant for confounding adjustment had *U* been measured and controlled for (Example 3.11).

A valid proxy for a confounding variable should (i) be associated with the hypothesized confounding factor *U* after controlling for exposure and (ii) not be associated with the outcome of interest except via *U*. Fig. 3.3 illustrates an example of causal associations required for a valid proxy, where *N* is now the proxy for *U* (Lipsitch et al., 2010).

Sometimes results are reported with stratification or restriction on a proxy variable in the form of subgroup analyses, in which strata were defined by a measured proxy variable. In other situations, results are reported with regression model adjustment for a proxy variable (sometimes results are reported with and without adjustment for a covariate that is a proxy for the confounder).

Example 3.12 illustrates the point that an analysis restricted to one level of a valid proxy variable (e.g. in which there is presumed to be little variation in the confounder *U*) might be viewed as less susceptible to confounding by *U*. However, as noted previously regarding residual confounding, the degree to which the proxy variable is a good surrogate for the unmeasured confounder will affect the degree by which confounding by *U* is minimized (Ogburn and VanderWeele, 2012; Ogburn et al., 2021). Moreover, the degree of residual bias that remains is typically proportional to the amount of confounding originally present (Greenland and Robins, 1985; Savitz and Barón, 1989).

3.3.3 Tool C-3: triangulation across studies

As described in Chapter 1, an *IARC Monographs* evaluation of an agent typically involves comparing findings across studies; this permits consideration of results across a set of studies that may differ in control for a confounder of concern within the wider context of the strengths and limitations of the available studies. The term *triangulation* is used to describe a variety of approaches in which analysts use different types of evidence from different study designs or types that have different identifying conditions; these approaches leverage variation between studies, focusing on settings in which biases vary across study types. Triangulation involves comparing results for a common effect from two or more studies that are

Example 3.11. Using a proxy variable to evaluate confounding in a cohort of Seventh Day Adventist adherents

A Working Group can evaluate concern about confounding by smoking if the reported results include analyses restricted to one level of a variable that is a proxy for smoking (the unmeasured potential confounder). An example is the study of chronic disease in the Adventist Health Study cohort, in which recruitment is restricted to a religious group who mostly do not smoke, to serve as a proxy for not smoking (Butler et al., 2008). (text continues on page 78)

Example 3.12. Restriction to one level of a proxy variable to examine residual confounding

Sheikh et al. (2020) examined the association between opium use (*E*) and oesophageal cancer (*D*) in the Islamic Republic of Iran; the Working Group discussed concerns about potential residual confounding by tobacco use (*U*). Sex was a measured variable in the study; it is presumed to be associated with tobacco smoking, because tobacco use is very rare among women in this population. A Working Group could consider sex as a proxy variable to indirectly assess residual confounding of the association between opium use and oesophageal cancer by smoking. In an analysis restricted to women, a positive association between opium use and oesophageal cancer was observed, and the association observed among women was similar in magnitude to that observed among men. Results conditioned on sex, if sex is considered a valid proxy for smoking, should be less susceptible to confounding by smoking. In this example, results suggested that the (sex- and smoking-adjusted) association between opium use and oesophageal cancer was unlikely to be substantially biased by residual confounding by tobacco smoking. (text continues on page 78)

thought to differ in susceptibility to confounding, or where the presumed confounder is thought to act in opposing directions; deliberate use could be made of studies conducted in contexts with differing confounding structures (Lawlor et al., 2016). Triangulation between covariate-adjusted analyses and instrumental variable analyses (such as Mendelian randomization studies) can offer some insight into whether the covariate-adjusted studies are likely to be confounded, because of the different identifying conditions required for covariate-adjusted analyses and Mendelian randomization studies (Example 3.13). Notably, there are also more advanced methods, such as multivariable Mendelian randomization, that adjust for known confounders to test the independence assumption in the Mendelian randomization studies (Brookhart et al., 2010; Burgess and Thompson, 2015). For further discussion of the use of triangulation in evidence synthesis, see Chapter 6.

Insight into possible bias can also be obtained by comparing results from two or more studies that are thought to differ in susceptibility to confounding. For example, a reviewer may raise a concern about a confounder that is uncontrolled in one or more studies (e.g. no control for smoking in studies among workers exposed to diesel fumes where lung cancer is the outcome) but observe that other studies of the same association reported similar results after adjustment for smoking (e.g. Bhatia et al., 1998). This offers another possible method to assess confounding; however, such simple comparisons across studies may not be valid. Rather, it would be surprising if the bias in one study applied perfectly to other studies (or even to other study samples drawn from the same source population). Confounding is seldom, if ever, the only bias of concern. When multiple biases are present, comparison between studies becomes more difficult (see Section 6.3 for further discussion and examples).

Example 3.13. Evidence triangulation to evaluate confounding

The *IARC Monographs* Volume 124 on night shift work found *limited* evidence that night shift work causes cancer in humans, with convincing evidence that it disrupts circadian rhythms (IARC, 2020). In that review, an example of triangulation between covariate-adjusted analyses and an instrumental variable analysis was discussed. A multivariable regression analysis demonstrated that, when examining chronotype (morning or evening preference) as a measure of circadian rhythm, morning preference was inversely associated with breast cancer incidence among participants in the UK Biobank study. The investigators identified genetic variants related to chronotype and undertook a Mendelian randomization study of chronotype and breast cancer incidence; they found a protective effect of morning preference on breast cancer risk (Richmond et al., 2019). This lends indirect support to the hypothesis that shift work is related to cancer risk because it disrupts this biological pathway. (text continues on page 79)

3.3.4 Tool C-4: bias adjustment

Investigators may be concerned about confounding by an unmeasured variable (the total or residual confounding) or the confounding produced by specific unmeasured variables. In the latter case, suppose that a reviewer has drawn a simple signed DAG for posited confounder–exposure and confounder–disease associations, implying potential bias in the study under review.

It is then necessary to assess how large this bias is likely to be, relative to the observed exposure–disease association. In assessing the potential impact of an unmeasured or incompletely adjusted confounder, reviewers may be able to estimate the size of the bias induced and decide whether it is indeed relevant. A variety of methods are available to quantitatively assess confounding under specified scenarios (or to identify bounds on bias due to an unmeasured confounder). Not all proposed methods are reviewed here; only a few approaches that are well suited to the *IARC Monographs* process are highlighted. Although subject matter knowledge is necessary, it need not be certain or complete; a range of values can be examined to assess plausible scenarios.

In the following subsections, many of the quantitative bias analyses are framed to guide judgement regarding whether a published estimate of association could plausibly be attributed entirely to an unmeasured confounder. As noted in Chapter 1, this reflects one of the primary questions posed to experts involved in an *IARC Monographs* review: can confounding reasonably be ruled out as an explanation for all of an observed exposure–cancer association? Simple expressions (and spreadsheet calculators) are also provided to facilitate the assessment of a range of bias. The focus throughout is on a single unmeasured confounder of primary concern; in Chapter 6, methods are extended to address multiple-bias analysis.

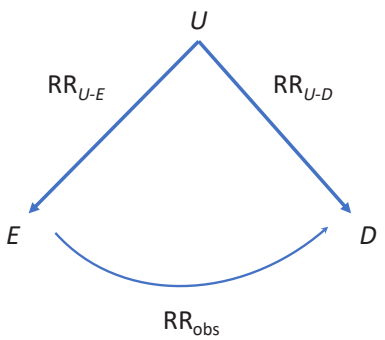
(a) Bounding

Concern about potential unmeasured confounders is often focused first on established cancer risk factors that have a strong independent association with the cancer of interest. This is because the understanding of strong risk factors for cancer outcomes is often better than that of the determinants of exposure. For simplicity, let us focus on settings where the hypothesized confounder increases the risk of cancer (i.e. $RR_{U\text{-}D} \geq 1$), where $RR_{U\text{-}D}$ denotes the magnitude of the confounder–outcome relative risk (this magnitude is typically estimated from prior information; Fig. 3.4).

Key message

If $RR_{U\text{-}D}$ is less than RR_{obs}, the reported relative risk between the exposure and the outcome in the study under review, then confounding by *U* cannot entirely explain an observed association.

Fig. 3.4. Diagram for analyses involving bounding and correction for the effect of a confounder.



Example 3.14. Use of bounding to examine confounding scenarios

Suppose that, in a study under review, it was reported that the observed association between opium use and laryngeal cancer (unadjusted for tobacco use) was RR_{obs} = 2.0. Suppose that confounding by tobacco smoking is of concern but had not been assessed in the study. On the basis of prior literature (Bakhshaee et al., 2017; Alizadeh et al., 2020), it can be hypothesized that the smoking–laryngeal cancer association in the study population was no larger than $RR_{U\text{-}D}$ = 5.5. In that situation, bounds on the smoking-adjusted association between opium use and laryngeal cancer are [0.36, 2.00]. With these assumptions, a reviewer could conclude that the observed association between opium use and laryngeal cancer could be due to confounding by smoking. (text continues below)

As shown in Example 3.14, if we know just the magnitude of the confounder–outcome association, $RR_{U\text{-}D}$, then, given a reported association,

$$RR_{obs} \times (1/RR_{U\text{-}D}) = \text{lower bound} \quad (3.1)$$

$$RR_{obs} \times 1 = \text{upper bound} \quad (3.2)$$

Key message

From these expressions, it follows that if the association between the confounder and outcome is small (i.e. $RR_{U\text{-}D}$ is close to 1) then the amount of uncontrolled bias from this confounder is also likely to be small.

RR_{obs}, it is possible to identify bounds (under a worst-case scenario, in which all the exposed have the confounder but none of the unexposed has the confounder) on the association of interest after adjustment for *U* (Flanders and Khoury, 1990):

Alternatively, if the magnitude of the confounder–exposure relative risk ($RR_{U\text{-}E}$) is less than RR_{obs}, then confounding by *U* cannot entirely explain an observed association. In other words, for confounding to entirely explain the observed association, both of the underlying associations ($RR_{U\text{-}E}$ and $RR_{U\text{-}D}$), not just one of them, must be larger than the published relative risk estimate, RR_{obs} (Cornfield et al., 1959). More informative bounds can be obtained using these two pieces of information ($RR_{U\text{-}E}$ and $RR_{U\text{-}D}$) (Flanders and Khoury, 1990; VanderWeele and Ding, 2017).

(b) Bias adjustment

As shown in Example 3.15, a simple bias-adjusted (Bross, 1966; Axelson, 1978; Schlesselman, 1978) estimate of the association can be derived, based on posited values for the strength of the confounder–outcome ($RR_{U\text{-}D}$) association and the prevalence of the confounder among the unexposed ($p_0 = \Pr[U = 1 \mid E = 0]$) and the exposed ($p_1 = \Pr[U = 1 \mid E = 1]$):

$$RR_{adj} = RR_{obs} \frac{RR_{U\text{-}D}\, p_0 + (1 - p_0)}{RR_{U\text{-}D}\, p_1 + (1 - p_1)} \quad (3.3)$$

Either unique values for p_0 and p_1 can be posited, along with the confounder–disease association ($RR_{U\text{-}D}$), or a range of plausible values for each can be posited and a distribution developed of the probable effects of bias due to an unmeasured confounder, using either Monte Carlo simulations or Bayesian priors (Steenland and Greenland, 2004). If the prevalences of the confounder among the unexposed and the exposed are not known, a Working Group member might take the latter approach to investigate what prevalence of smoking would be needed to entirely explain the observed association, and then consider the plausibility of such a pattern in the study population. Implementation of such calculations in a spreadsheet facilitates exploration (Fox et al., 2021).

(c) Unknown uncontrolled confounders and E*-values*

If a concern is expressed about an unknown confounder, a reviewer might undertake a quantitative bounding analysis, following the principles outlined previously in Section 3.3.4(a). Such an evaluation could be considered when doubts remain about causality, despite the lack of an identified confounder. For example, if it is arbitrarily assumed that the magnitudes of the associations of the confounder with exposure and outcome are equal on a risk-ratio scale (i.e. $RR_{U\text{-}E} = RR_{U\text{-}D}$) then, for an observed positive exposure–outcome association to be entirely due to a confounder *U*, $RR_{U\text{-}E}$

Example 3.15. Bias adjustment to evaluate confounding

Consider the possibility of unmeasured smoking as a potential confounder in a study of opium use and lung cancer. Suppose that the prevalence of smoking in the unexposed is 20%, the prevalence in the exposed is 30% ($RR_{U\text{-}E}$ = 1.5), smoking has a hypothesized $RR_{U\text{-}D}$ of 10, and the observed relative risk for opium and lung cancer is 2.0 (exposed versus unexposed). Let p_0 be the proportion of smokers among the unexposed and p_1 be the proportion of smokers among the exposed. The risk of lung cancer among those unexposed due solely to smoking will be a weighted average of the risks of lung cancer in non-smokers and smokers, i.e. $RR_{U\text{-}D}\ p_0 + (1 - p_0)$, and the risk of lung cancer among the exposed, due to smoking alone, is $RR_{U\text{-}D}\ p_1 + (1 - p_1)$. The relative risk of exposed versus unexposed, due to smoking alone, is $[RR_{U\text{-}D}\ p_1 + (1 - p_1)]/[RR_{U\text{-}D}\ p_0 + (1 - p_0)]$, and we can adjust the observed relative risk due to opium by this factor to indirectly adjust for the estimated confounding by smoking (Flanders and Khoury, 1990). (text continues on page 81)

$$\mathrm{RR_{adj}} = \mathrm{RR_{obs}}\frac{\mathrm{RR}_{U-D}p_0 + (1-p_0)}{\mathrm{RR}_{U-D}p_1 + (1-p_1)} = \mathrm{RR_{obs}}\frac{10(0.2) + (1-0.2)}{10(0.3) + (1-0.3)} = \mathrm{RR_{obs}}\frac{2.8}{3.7} \quad \text{(E3.1)}$$

For a simple bias adjustment, as given by this equation, one can correct the observed risk ratio for the potential confounding; if the observed risk ratio were 2.00, the adjusted risk ratio would be

$$\frac{2.8}{3.7} \times 2.0 = 1.5 \quad \text{(E3.2)}$$

and $RR_{U\text{-}D}$ must equal, or exceed, RR_{obs} + sqrt[RR_{obs} × (RR_{obs} − 1)], a quantity that has been termed the *E*-value (VanderWeele and Ding, 2017). Note that this value is derived using just the observed (potentially confounded) association between agent and outcome, RR_{obs}, without specification of the confounder–outcome or confounder–exposure association (other than assuming that they are equal). It also unrealistically assumes that the prevalence of the uncontrolled confounder among the exposed is 100% or, equivalently, that the prevalence of the exposure among those without the confounder is 0%, and hence can be misleadingly small compared with what is needed for an actual confounder to fully explain the magnitude of RR_{obs} (MacLehose et al., 2021), as shown in Example 3.16.

Bias analyses (Flanders and Khoury, 1990; Lash et al., 2009; Fox et al., 2021; MacLehose et al., 2021) allow one to relax the assumptions used by the *E*-value that $RR_{U\text{-}E}$ equals $RR_{U\text{-}D}$ and that the prevalence of the confounder is 100% among the exposed.

3.4 Summary

Confounding is typically of concern in observational studies. Expert reviewers can assess the impact of confounding on the observed exposure–cancer association in several ways. Some study designs can minimize confounding, for example by matching on probable confounders ahead of time. In other studies, the investigators will have measured potential confounders and controlled for them in the design or analysis.

Key message

When there is concern about unknown confounders, a quantitative bounding analysis, as discussed previously in Section 3.3.4(a), can clarify what magnitudes of confounder–disease association, and what prevalences of confounder among exposed and unexposed, would be needed to entirely explain an observed exposure–disease association (see Side Box 3.5).

Example 3.16. The *E*-value to evaluate confounding

The reported association between opium use and oesophageal cancer (unadjusted for some unknown confounder *U*) was RR_{obs} = 2.0 (Example 3.15). Suppose that confounding by the unknown confounder *U* is suspected. The resultant *E*-value would take a value of 2 + sqrt[2 × (2 − 1)] = 3.4, meaning that if a reviewer posited that $RR_{U\text{-}E}$ and $RR_{U\text{-}D}$ were positive, equal, and both less than 3.4, it could be concluded that confounding by *U* could not entirely explain the observed positive exposure–disease association.

However, a reviewer might assume the confounder–opium use association, $RR_{U\text{-}E}$, to be larger than 3.4. This illustrates one important caution concerning interpretation of the *E*-value: although it might be tempting to say that both associations need to be at least as large as the *E*-value, that is incorrect. In fact, $RR_{U\text{-}D}$ could be less than the *E*-value, while $RR_{U\text{-}E}$ could be substantially larger than the *E*-value, allowing for confounding to completely explain the association. Conversely, both $RR_{U\text{-}D}$ and $RR_{U\text{-}E}$ could be substantially larger than 3.4 and still not completely explain the association, for the simple reason that the unknown confounder *U* could have a prevalence substantially less than 100% among opium users. (text continues on page 82)

Side Box 3.5. Information needed to facilitate use of bias assessment to evaluate confounding

For bounding approaches, the original (or associated) studies should report the value of the probable magnitude of association of the confounder with the outcome of interest in the population under study, and the association of the confounder with the exposure of interest.

For quantitative bias assessment, the original or associated studies should also report, more specifically, the prevalence of the confounder among those unexposed (p_0) and exposed (p_1) to the agent of interest. (text continues on page 82)

In both settings, reviewers will want to consider whether the confounder was well measured and controlled (i.e. whether residual confounding is likely to remain). Reviewers may also consider whether, based on the literature, there are likely to be important unmeasured confounders.

If potential confounders were not measured or were inadequately controlled in a study, then reviewers need to make informed judgements about the direction of residual confounding and its probable magnitude, and, in particular, the extent to which residual confounding could explain the observed exposure–disease association. The reliability of such judgements will be greatly improved to the extent that they make use of background information about the relations of uncontrolled potential confounders to the exposure and disease under study, and the results of other studies that did control for those potential confounders.

References

Alizadeh H, Naghibzadeh Tahami A, Khanjani N, Yazdi-Feyzabadi V, Eslami H, Borhaninejad V, et al. (2020). Opium use and head and neck cancers: a matched case-control study in Iran. Asian Pac J Cancer Prev. 21(3):783–90. doi:10.31557/APJCP.2020.21.3.783 PMID:32212808

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature. 526(7571):68–74. doi:10.1038/nature15393 PMID:26432245

Axelson O (1978). Aspects on confounding in occupational health epidemiology. Scand J Work Environ Health. 4(1):98–102. doi:10.5271/sjweh.2720 PMID:644270

Bakhshaee M, Raziee HR, Afshari R, Amali A, Roopoosh M, Lotfizadeh A (2017). Opium addiction and risk of laryngeal and esophageal carcinoma. Iran J Otorhinolaryngol. 29(90):19–22. PMID:28229058

Bhatia R, Lopipero P, Smith AH (1998). Diesel exhaust exposure and lung cancer. Epidemiology. 9(1):84–91. doi:10.1097/00001648-199801000-00017 PMID:9430274

Brookhart MA, Rassen JA, Schneeweiss S (2010). Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 19(6):537–54. doi:10.1002/pds.1908 PMID:20354968

Bross ID (1966). Spurious effects from an extraneous variable. J Chronic Dis. 19(6):637–47. doi:10.1016/0021-9681(66)90062-2 PMID:5966011

Buckley JP, Keil AP, McGrath LJ, Edwards JK (2015). Evolving methods for inference in the presence of healthy worker survivor bias. Epidemiology. 26(2):204–12. doi:10.1097/EDE.0000000000000217 PMID:25536456

Burgess S, Thompson SG (2015). Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 181(4):251–60. doi:10.1093/aje/kwu283 PMID:25632051

Butler TL, Fraser GE, Beeson WL, Knutsen SF, Herring RP, Chan J, et al. (2008). Cohort profile: the Adventist Health Study-2 (AHS-2). Int J Epidemiol. 37(2):260–5. doi:10.1093/ije/dym165 PMID:17726038

Chan DS, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, et al. (2011). Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. PLoS One. 6(6):e20456. doi:10.1371/journal.pone.0020456 PMID:21674008

Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. (2010). Illustrating bias due to conditioning on a collider. Int J Epidemiol. 39(2):417–20. doi:10.1093/ije/dyp334 PMID:19926667

Cook NR, Cole SR, Hennekens CH (2002). Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. Am J Epidemiol. 155(11):1045–53. doi:10.1093/aje/155.11.1045 PMID:12034583

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst. 22(1):173–203. PMID:13621204

Davey Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S (2007). Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Med. 4(12):e352. doi:10.1371/journal.pmed.0040352 PMID:18076282

Flanders WD, Khoury MJ (1990). Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. Epidemiology. 1(3):239–46. doi:10.1097/00001648-199005000-00010 PMID:2081259

Fox MP, MacLehose RF, Lash TL (2021). Misclassification. In: Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer; pp. 141–195. doi:10.1007/978-3-030-82673-4_6

Frisell T, Öberg S, Kuja-Halkola R, Sjölander A (2012). Sibling comparison designs: bias from non-shared confounders and measurement error. Epidemiology. 23(5):713–20. doi:10.1097/EDE.0b013e31825fa230 PMID:22781362

Greenland S (1980). The effect of misclassification in the presence of covariates. Am J Epidemiol. 112(4):564–9. doi:10.1093/oxfordjournals.aje.a113025 PMID:7424903

Greenland S (2012). Intuitions, simulations, theorems: the role and limits of methodology. Epidemiology. 23(3):440–2. doi:10.1097/EDE.0b013e31824e278d PMID:22475828

Greenland S, Mansournia MA, Altman DG (2016). Sparse data bias: a problem hiding in plain sight. BMJ. 352:i1981. doi:10.1136/bmj.i1981 PMID:27121591

Greenland S, Pearl J, Robins JM (1999b). Causal diagrams for epidemiologic research. Epidemiology. 10(1):37–48. doi:10.1097/00001648-199901000-00008 PMID:9888278

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

Greenland S, Robins JM, Pearl J (1999a). Confounding and collapsibility in causal inference. Stat Sci. 14(1):29–46. doi:10.1214/ss/1009211805

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, Robins JM (2006). Estimating causal effects from epidemiological data. J Epidemiol Community Health. 60(7):578–86. doi:10.1136/jech.2004.029496 PMID:16790829

Hernán MS, Robins JM (2023). Causal inference: what if. Boca Raton (FL), USA: Chapman and Hall/CRC. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

Hertz-Picciotto I (2000). Invited commentary: shifting the burden of proof regarding biases and low-magnitude associations. Am J Epidemiol. 151(10):946–8, discussion 949–50. doi:10.1093/oxfordjournals.aje.a010136 PMID:10853632

IARC (2018). Red meat and processed meat. IARC Monogr Eval Carcinog Risks Hum. 114:1–506. Available from: https://publications.iarc.who.int/564 PMID:29949327

IARC (2019). Preamble to the *IARC Monographs* (amended January 2019). Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

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

Larjavaara S, Schüz J, Swerdlow A, Feychting M, Johansen C, Lagorio S, et al. (2011). Location of gliomas in relation to mobile telephone use: a case-case and case-specular analysis. Am J Epidemiol. 174(1):2–11. doi:10.1093/aje/kwr071 PMID:21610117

Lash TL, Fink AK, Fox MP (2009). Multiple bias modeling. In: Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. New York (NY), USA: Springer; pp. 151–73. doi:10.1007/978-0-387-87959-8_9

Lawlor DA, Tilling K, Davey Smith G (2016). Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866–86. doi:10.1093/ije/dyw314 PMID:28108528

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

MacLehose RF, Ahern TP, Lash TL, Poole C, Greenland S (2021). The importance of making assumptions in bias analysis. Epidemiology. 32(5):617–24. doi:10.1097/EDE.0000000000001381 PMID:34224472

Maclure M (1998). The case-specular study design and counterfactual controls. Epidemiology. 9(1):6–7. doi:10.1097/00001648-199801000-00003 PMID:9430260

Mansournia MA, Jewell NP, Greenland S (2018). Case–control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 33(1):5–14. doi:10.1007/s10654-017-0325-0 PMID:29101596

McClure ES, Vasudevan P, DeBono N, Robinson WR, Marshall SW, Richardson D (2020). Cancer and noncancer mortality among aluminum smelting workers in Badin, North Carolina. Am J Ind Med. 63(9):755–65. doi:10.1002/ajim.23150 PMID:32649003

Naimi AI, Cole SR, Hudgens MG, Brookhart MA, Richardson DB (2013). Assessing the component associations of the healthy worker survivor bias: occupational asbestos exposure and lung cancer mortality. Ann Epidemiol. 23(6):334–41. doi:10.1016/j.annepidem.2013.03.013 PMID:23683709

Ogburn EL, Rudolph KE, Morello-Frosch R, Khan A, Casey JA (2021). A warning about using predicted values from regression models for epidemiologic inquiry. Am J Epidemiol. 190(6):1142–7. doi:10.1093/aje/kwaa282 PMID:33350434

Ogburn EL, VanderWeele TJ (2012). On the nondifferential misclassification of a binary confounder. Epidemiology. 23(3):433–9. doi:10.1097/EDE.0b013e31824d1f63 PMID:22450692

Pearce NE, Smith AH, Howard JK, Sheppard RA, Giles HJ, Teague CA (1986). Non-Hodgkin's lymphoma and exposure to phenoxyherbicides, chlorophenols, fencing work, and meat works employment: a case-control study. Br J Ind Med. 43(2):75–83. doi:10.1136/oem.43.2.75 PMID:3753879

Pierce BL, Kraft P, Zhang C (2018). Mendelian randomization studies of cancer risk: a literature review. Curr Epidemiol Rep. 5(2):184–96. doi:10.1007/s40471-018-0144-1 PMID:30034993

Pinkerton LE, Hein MJ, Anderson JL, Little MP, Sigurdson AJ, Schubauer-Berigan MK (2016). Breast cancer incidence among female flight attendants: exposure–response analyses. Scand J Work Environ Health. 42(6):538–46. doi:10.5271/sjweh.3586 PMID:27551752

Richmond RC, Anderson EL, Dashti HS, Jones SE, Lane JM, Strand LB, et al. (2019). Investigating causal relations between sleep traits and risk of breast cancer in women: mendelian randomisation study. BMJ. 365:l2327. doi:10.1136/bmj.l2327 PMID:31243001

Robins J (1986). A new approach to casual inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. Math Model. 7(9–12):1393–512. doi:10.1016/0270-0255(86)90088-6

Savitz DA, Barón AE (1989). Estimating and correcting for confounder misclassification. Am J Epidemiol. 129(5):1062–71. doi:10.1093/oxfordjournals.aje.a115210 PMID:2705426

Schisterman EF, Cole SR, Platt RW (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 20(4):488–95. doi:10.1097/EDE.0b013e3181a819a1 PMID:19525685

Schlesselman JJ (1978). Assessing effects of confounding variables. Am J Epidemiol. 108(1):3–8. PMID:685974

Sheikh M, Shakeri R, Poustchi H, Pourshams A, Etemadi A, Islami F, et al. (2020). Opium use and subsequent incidence of cancer: results from the Golestan Cohort Study. Lancet Glob Health. 8(5):e649–60. doi:10.1016/S2214-109X(20)30059-0 PMID:32353313

Sjölander A, Öberg S, Frisell T (2022). Generalizability and effect measure modification in sibling comparison studies. Eur J Epidemiol. 37(5):461–76. doi:10.1007/s10654-022-00844-x PMID:35312926

Smith GD, Ebrahim S (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 32(1):1–22. doi:10.1093/ije/dyg070 PMID:12689998

Steenland K, Greenland S (2004). Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol. 160(4):384–92. doi:10.1093/aje/kwh211 PMID:15286024

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

VanderWeele TJ (2016). Mediation analysis: a practitioner's guide. Annu Rev Public Health. 37(1):17–32. doi:10.1146/annurev-publhealth-032315-021402 PMID:26653405

VanderWeele TJ, Ding P (2017). Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 167(4):268–74. doi:10.7326/M16-2607 PMID:28693043

Wade KH, Yarmolinsky J, Giovannucci E, Lewis SJ, Millwood IY, Munafò MR, et al.; with the MR in Nutrition, Cancer working group (2022). Applying Mendelian randomization to appraise causality in relationships between nutrition and cancer. Cancer Causes Control. 33(5):631–52. doi:10.1007/s10552-022-01562-1 PMID:35274198

Yarmolinsky J, Wade KH, Richmond RC, Langdon RJ, Bull CJ, Tilling KM, et al. (2018). Causal inference in cancer epidemiology: what is the role of Mendelian randomization? Cancer Epidemiol Biomarkers Prev. 27(9):995–1010. doi:10.1158/1055-9965.EPI-17-1177 PMID:29941659

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

4.1 Introduction 88
4.2 Qualitative evaluation of the direction of bias due to errors in exposures 90
4.3 Tools for quantifying bias due to errors in exposure 100
4.4 Outcome misclassification 110
4.5 Summary 116

CHAPTER 4.

Information bias: misclassification and mismeasurement of exposure and outcome

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

4.1 Introduction

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

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

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

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

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

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

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

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

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

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

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

4.2.1 Non-differential errors in exposure

(a) Measurement errors of continuous variables

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

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

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

Key message

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

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

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

(b) Misclassification of categorical variables

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

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

Key message

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

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

Key message

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

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

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

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

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

(a) Classical model

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

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

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

(b) Linear model

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

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

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

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

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

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

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

(c) Berkson model

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

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

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

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

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

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

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

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

Key message

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

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

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

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

4.2.3 Differential errors in exposure

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

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

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

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

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

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

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

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

Key message

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

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

(a) Recall and information bias

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

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

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

(b) Differential information when provided by proxies

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

Example 4.7. Recall bias and knowledge of carcinogenicity

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

Example 4.8. Estimation of the extent of recall bias

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

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

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

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

4.2.4 Tools for assessing differential exposure information bias

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

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

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

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

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

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

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

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

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

Key message

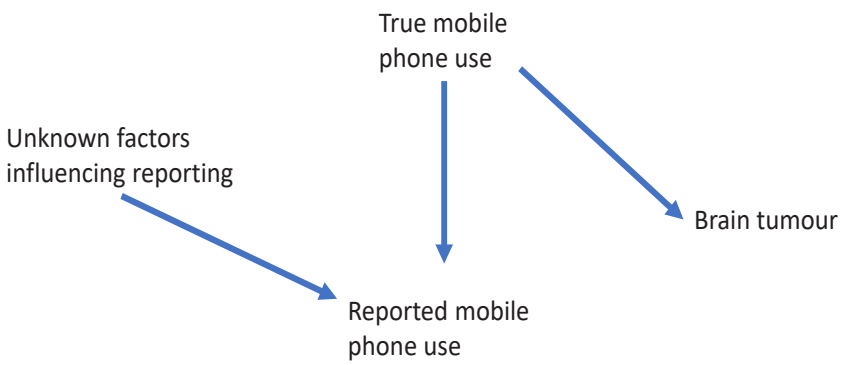
A key assumption of the use of negative control exposures is that any tendency for reduced or exaggerated recall of exposure is likely to be similar for the main study exposure and the negative control exposure.

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

Example 4.10. Using DAGs to identify recall bias

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

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



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

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

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

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

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

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

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

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

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

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

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

Example 4.13. Using national statistics to assess recall bias

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

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

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

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

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

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

Example 4.15. Using triangulation to assess recall bias

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

4.3 Tools for quantifying bias due to errors in exposure

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

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

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

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

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

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

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

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

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

4.3.2 Tool E-8: multidimensional analysis

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

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

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

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

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

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

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

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

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

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

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

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

Example 4.17. Analysis of bias from differential exposure misclassification

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

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

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

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

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

Example 4.18. Multidimensional sensitivity analysis

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

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

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

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

4.3.3 Limitations of methods for analyses of exposure measurement errors

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

4.3.4 Tool E-9: multiple categorical bias analysis

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

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

4.3.5 Tool E-10: probabilistic bias analysis

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

Example 4.19. Sensitivity analysis for both confounding and misclassification

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

Example 4.20. Sensitivity analysis for categorical exposure misclassification

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

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

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

CI, confidence interval; OR, odds ratio.

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

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

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

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

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

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

(a) Continuous measures of exposure

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

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

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

Example 4.21. Probabilistic bias analysis for exposure misclassification

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

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

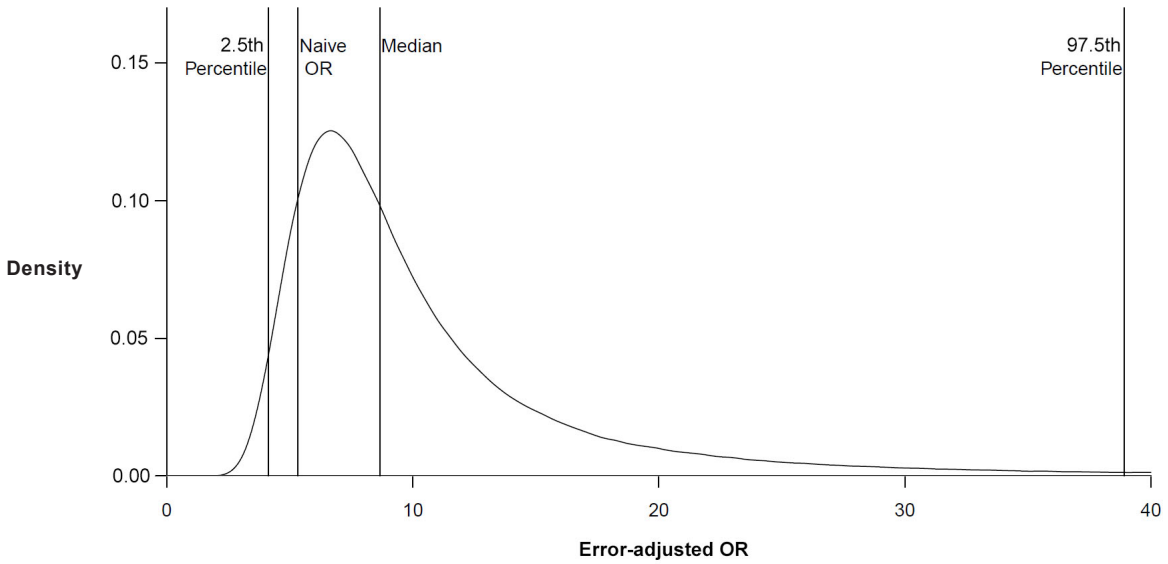
Bias parameter	Distribution parameters of triangular distribution		
	Minimum (%)	Mode (%)	Maximum (%)
se_1	66	79	89
sp_1	75	83	90
se_0	50	68	82
sp_0	87	93	96

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

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

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

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



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

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

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

Key message

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

Example 4.22a. Regression calibration for adjustment for measurement error

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and the upper confidence interval is given by

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

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

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

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

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

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

Key message

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

Example 4.23 illustrates this type of adjustment.

4.3.7 Tool E-12: other methods for quantifying bias

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

4.4 Outcome misclassification

4.4.1 Non-differential outcome misclassification

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

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

Example 4.23. Bias adjustment for misclassified categorical exposures

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

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

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

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

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

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

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

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

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

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

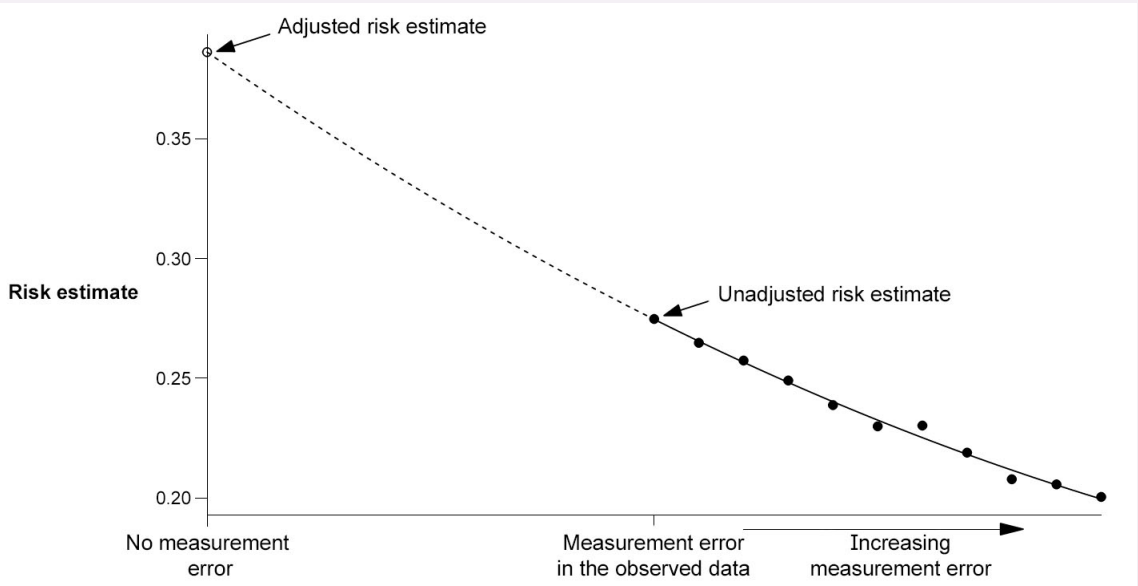
Method	Process description	Preferable in the following situations
Probabilistic bias analysis	Bias parameters are simulated.	Original study data are unavailable Bias model is known
MC-SIMEX	Increasing misclassification is simulated.	Exposure variable with more than two categories Multiple regression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

MC-SIMEX, simulation extrapolation for misclassification.

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

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

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



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

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

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

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

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

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

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

Key message

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

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

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

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

4.4.2 Differential outcome misclassification

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

4.4.3 Quantitative assessment of bias due to outcome misclassification

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

Side Box 4.3. Bayesian methods for error adjustment

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

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

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

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

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

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

Observed data

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

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

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

Side Box 4.4. Multiple imputation

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

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

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

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

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

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

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

Side Box 4.4. Multiple imputation (continued)

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

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

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

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

and control participants,

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

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

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

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

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

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

Example 4.28. Non-differential outcome misclassification of tumour subtypes

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

Example 4.29. Differential outcome misclassification among firefighters

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

Example 4.30. Sensitivity analysis for outcome misclassification

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

4.5 Summary

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

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

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

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

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

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 5. Selection bias and other miscellaneous biases

5.1 Introduction 124
5.2 Identifying selection bias in cohort studies 125
5.3 Identifying selection bias in case–control studies 132
5.4 Tools for assessing and adjusting for selection bias 139
5.5 Other miscellaneous biases 149
5.6 Summary 154

CHAPTER 5.

Selection bias and other miscellaneous biases

Neil Pearce, Laura Beane Freeman, Manolis Kogevinas, Richard MacLehose, Ellen Aagaard Nøhr, Marie-Elise Parent, and Lorenzo Richiardi

5.1 Introduction

Epidemiological studies are intended to obtain valid exposure effect estimates for a target population (e.g. all women aged 20 years or older). In practice, specific epidemiological studies are, or at least should be, based on a clearly defined source population (e.g. all women in France aged 20 years or older), followed up over a clearly defined risk period (e.g. 2010–2020). However, it is rare for a study to include all of the source population over the entire risk period. In cohort studies, there could be incomplete recruitment at baseline, and some participants may be lost to follow-up. Case–control studies, by design, involve recruiting a sample of control participants from the source population, and there may also be incomplete recruitment of case or control participants; this may create selection bias if, as a result, the two groups are different from the full source population with respect to exposure status or level.

Selection bias is present when the effect estimate (e.g. the odds ratio [OR]) of the association between the exposure and the outcome in the study population is different from that in the source population, because of selective recruitment into the study or selective loss to follow-up. Thus, the defining characteristic of selection bias is that it occurs as a result of differences between the study population and the source population from which it is selected. Selection bias can occur for a variety of reasons, either during initial recruitment from the source population (e.g. differential recruitment with respect to both the exposure and the outcome) or during follow-up (e.g. differential retention in the study). In a published paper, selection bias can be particularly difficult to assess, for example by *IARC Monographs* Working Groups, because few papers report the information required to assess and quantify it.

Selection bias is distinct from issues of generalizability (or transportability) (Richiardi et al., 2013). The terms *representativeness*, *generalizability*, and *transportability* refer to comparisons between the target population and the source population. In most studies, the concept of the target population is left undefined, and there is no need to invoke some hypothetical target population to validly design and analyse a study. Moreover, if an exposure has a non-null effect in a defined source population, or even in a specific study population, this is of concern in

itself, irrespective of issues of transportability. Thus, in theory, issues of transportability are usually not central to *IARC Monographs* reviewers, because the focus is generally on whether there is a non-null effect in any population, rather than the size of the effect in a specific population. In contrast, evidence synthesis often does involve an assessment of consistency of results across studies, at least in qualitative rather than quantitative terms, and any major inconsistencies will require further consideration and explanation.

Selection bias is often confused with issues of representativeness (Munafò et al., 2018) but these are very different concepts (see Chapter 2). In fact, many important causal associations (e.g. smoking and lung cancer) have been discovered or confirmed in studies involving particular subgroups of the general population, such as the classic study of smoking and lung cancer in British doctors (Hill and Doll, 1956). Thus, a study should not be assumed to suffer from selection bias simply because it is not based on a random sample of the general population.

According to this definition of selection bias, if information is obtained for all of the source population over the entire risk period, then the study population is the same as the source population; therefore, selection bias does not occur. Defined in this way, selection bias closely aligns with collider bias (see Chapter 2; Hernán et al., 2004; Pearce and Richiardi, 2014), arising because it is only possible to analyse data for those who have been included in the study, and therefore the analysis is conditioned on selection into the study. Selection bias is not only the result of collider stratification. It can also occur when selection is associated with effect modifiers. Without stratification by, or standardization over, those modifiers, the effect estimated in such a study may be very different from the effect that would have been estimated in the source population. This type of selection bias may be less relevant in the context of cancer hazard identification. Example 5.1 examines selection bias in a case–control study.

Key message

In general, important selection bias will occur if the selection (through either recruitment or loss to follow-up) is associated with both exposure and disease status together (e.g. if exposed case participants are more likely or less likely than other groups to be recruited) (Richiardi et al., 2013). Therefore, this chapter focuses on the situation in which selection is associated with both exposure and disease.

A primary question posed to expert reviewers, such as *IARC Monographs* Working Groups, in the context of hazard identification is, "Can we reasonably rule out selection bias as an explanation for an observed exposure–cancer association?" This can be particularly difficult to assess, because most published studies provide little or no discussion of the potential for selection bias, in contrast to the usually more extensive discussions of the potential for confounding (Chapter 3) or misclassification (Chapter 4).

Key message

Although selection bias is often the most mathematically simple bias for which estimates of effect (see Chapter 7) can be bias-adjusted, the information needed for such bias adjustments is rarely available or reported in published papers.

One exception to the typical lack of available information is the literature on the Interphone study (Cardis et al., 2010); this example is used frequently in this chapter, although it is recognized that this level and detail of information is usually not available to *IARC Monographs* Working Groups or other expert reviewers.

This chapter starts by discussing selection bias in cohort studies and then considers the additional forms of selection bias that can occur in case–control studies. Methods are then presented for assessing selection bias in a published paper.

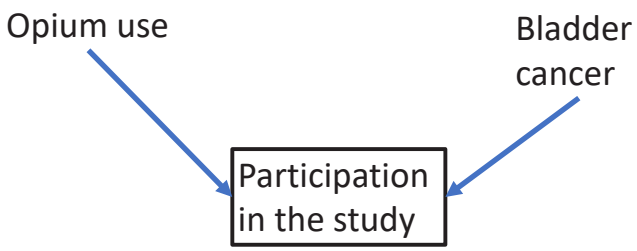
5.2 Identifying selection bias in cohort studies

Many cohort studies of cancer rely on the willingness of people to participate, both at baseline and during follow-up. Furthermore, the researchers may choose different inclusion and restriction strategies in specific analyses that may also affect the composition of the study population. When reviewing such studies, it is important to consider whether such selection may have biased the results. The relation of selection bias to other types of bias is defined in Chapter 2.

Example 5.1. Selection bias in a case–control study

Fig. 5.1 illustrates the occurrence of selection bias in a case–control study of opium use and bladder cancer, an example discussed in more detail later in this chapter. Chapter 2 introduced the use of directed acyclic graphs (DAGs) to ascertain the possible presence of selection bias. In this DAG, participation in the study is affected by both opium use and bladder cancer. Opium users may be more hesitant than non-opium users to participate in a study. People with bladder cancer (potential case participants) may be more likely than control subjects to participate in the study (e.g. because of their interest in the subject) or less likely to participate because of their illness status. A box is drawn around "Participation in the study" to indicate that all analyses condition on this factor (i.e. analyses are limited to this group). In this example, an opium use–bladder cancer association could be found in a study, even if one did not truly exist. Alternatively, if this were not a case–control study but, rather, included the entire source population, then selection bias could not occur, because everyone would be enrolled in the study and nothing could affect participation. Thus, whether someone had used opium or had bladder cancer would not affect participation in the study, and there would be no arrow from "Opium use" or " "Bladder cancer" to "Participation in the study". (text continues on page 125)

Fig. 5.1. Illustrative example of possible selection bias in a case–control study of opium use and bladder cancer.



5.2.1 Non-response at baseline

The first stage at which selection bias may occur in a cohort study is in the initial recruitment into a study. As discussed in the previous section, if the entire source population is recruited, which may be the situation in a register-based study that does not rely on consent to participate, then selection bias cannot occur (at least at baseline). However, even if there is incomplete recruitment or participation, the study population can still provide unbiased effect estimates (or, at least, estimates that are unbiased by selection issues). For example, a study with a 40% response rate at baseline may nevertheless be almost completely unbiased if non-response is not associated with either exposure or disease. Selection at baseline that is related to a particular exposure (e.g. socioeconomic status [SES]) should not bias future results, as long as participation is not also associated with future disease status (e.g. if affluent people are more likely to participate than non-affluent people, but their participation is not related to whether they will or will not develop the disease being studied). However, if exposure and outcome jointly determine selection (e.g. affluent people who will eventually develop the disease are more likely to participate in the study, or non-affluent people who will stay healthy are more likely to participate than others), this will result in a selection bias arising because the analysis includes only those who participated in the study (i.e. the analysis conditions on participation in the study) (see Chapter 2 and Section 5.1). There is also the possibility of selection bias if, instead of the outcome itself, it is an outcome risk factor that determines selection at initial recruitment, because that risk factor could alter the causal effect estimate in the study population, acting in the same way as a confounder (see Chapter 2).

However, such bias will usually be small, as shown in Example 5.2.

Side Box 5.1 outlines the key information that should be reported to facilitate assessment of bias due to non-response at baseline.

Traditionally, in cohort studies, the assumption has been that because potential participants are not aware of their risk of future disease at baseline, this will not influence their decision to participate, and selection at baseline has been considered a minor problem compared with loss to follow-up, which may be jointly determined by exposure and outcome. However, this has been questioned in the UK Biobank study, for which the initial response rate was only 5.5%, and in which it was shown that participation in the study was related to some particular exposures and outcomes (Fry et al., 2017; Munafò et al., 2018). However, this bias would apply only to the cross-sectional analyses of the baseline data, and will usually be small (see Example 5.2).

Moreover, Richiardi et al. (2013) have argued that this type of selection bias will not occur in a cohort study if people with prevalent disease at baseline (or who are diagnosed soon after baseline) are excluded, assuming that other factors that influence participation do not also affect disease (see Example 5.3). This is possible in cohorts for which electronic health record-linked data are available; this would enable the identification of cases of disease that occur after recruitment. Therefore, for *IARC Monographs* Working Groups it is important to consider the probable latency period (usually assumed to be about 5 years for cancer) during which disease may be present but not yet diagnosed.

5.2.2 Loss to follow-up

Selection bias may also occur when loss to follow-up differs between exposed and unexposed people, because this is related to the ability to observe disease outcomes.

Key message

Selection bias occurring from loss to follow-up is perhaps of more concern than selection bias from recruitment in cohort studies (and in case–control studies based on them), because exposure, predictors of the outcome, and the outcome itself may now jointly determine participation.

Example 5.2. Magnitude of selection bias

Pizzi et al. (2011) demonstrated that when both the exposure and another risk factor that is independent from the exposure double the probability of selection into the study and the other risk factor also doubles the risk of the outcome, this selection bias will result in an observed relative risk of only 1.02 for the exposure–outcome association when the true relative risk is 1.0. Moreover, this bias can be corrected if the analyses are adjusted for the risk factors that determine the selection. In this example, it is assumed that the exposure is not associated with the other risk factor in the source population; if they were associated, the bias would be larger. (text continues above)

Side Box 5.1. Information that should be reported to enable the assessment of bias due to non-response at baseline

The key parameters that should be reported to enable the post-publication assessment of selection bias are the probability of participation in the study stratified on exposure and disease status. Unfortunately, these are rarely, if ever, available. In particular, for studies involving consent from the participants, this information will rarely be available for those who do not consent, although some information may be available from the sampling frame (e.g. some population registers include information on age and sex). Authors should report not only the overall response rate but also the response rates in key subgroups of interest by baseline exposure status. In addition, descriptive tables of participants and non-participants (with sex, SES, age, ethnic group, and major risk factors if possible) should be provided. (text continues above)

Example 5.3. Assessing bias due to non-response at baseline in an occupational cohort study of flight attendants

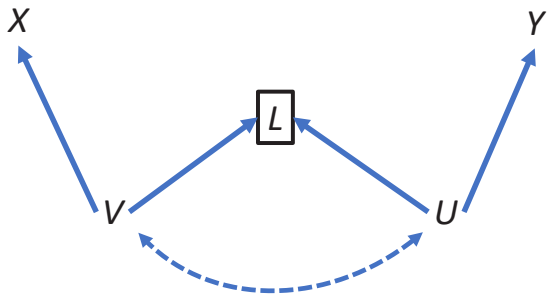
A study was conducted to evaluate the association of exposure to cosmic radiation and circadian disruptors with breast cancer risk in former flight attendants (Schubauer-Berigan et al., 2015). The response rate for inclusion in this study was 64.4%. Selection bias could have occurred if participation was related to employment characteristics as a flight attendant and also to the disease. The breast cancer incidence cohort of flight attendants was a subset of a cohort (the mortality cohort) of former flight attendants employed by Pan American World Airways (Pan Am) for at least 1 year, for which the main outcome considered was breast cancer mortality. The incidence cohort was assembled from the personnel records of Pan Am. Women (n = 9461) in the mortality cohort were invited to participate in the incidence cohort by completing a detailed telephone interview or mailed questionnaire (2002–2005), which contained questions about their demographic information, work history, and non-occupational risk factors for breast cancer (e.g. reproductive history and use of alcohol, tobacco, and hormone replacement therapy [HRT]). The next of kin of deceased flight attendants were also contacted and were each invited to complete the questionnaire about the decedent. Duration of employment was closely correlated with estimated cumulative exposure to cosmic radiation.

After some minor exclusions, the incidence cohort included all the respondents to the telephone interview and mailed questionnaire (n = 6093 women, 64.4% of the 9461 eligible women in the mortality cohort); 2% of the cohort overall and 8% of those with breast cancer were deceased. The response rate for proxies of decedents (n = 134) was lower (41%) than among living cohort members (65%). For women who died after a breast cancer diagnosis, the response rate was similarly low (46%). The median duration of Pan Am employment based on workplace records among the respondents was 5.8 years and was slightly longer than for the mortality cohort (5.0 years), suggesting that long-term employees of Pan Am were more likely to respond to the questionnaire. Other major sociodemographic differences between participants and non-participants were very small (Pinkerton et al., 2016). Although there were some small differences in response rates between deceased and living cohort members, the overall potential selection with regard to breast cancer could be expected to be minimal, given the small number of decedents and the lack of major differences in major socioeconomic and exposure factors between participants and non-participants. The fact that the participants had worked slightly longer than the non-participants, and therefore had more shift work (which is a potential circadian disruptor and thus a possible risk factor for breast cancer), is unlikely to have resulted in large selection bias, unless breast cancer risk also affected participation in the study. (text continues on page 127)

Selection bias from loss to follow-up can also occur in randomized trials, even when the exposure (which in a randomized trial would be the intervention) has been randomized at baseline. In particular, if there is loss to follow-up and this is jointly associated with both exposure status and outcome status, then selection bias can result, because all analyses will include only those participants for whom there are follow-up data. When both the exposure and the outcome affect participation in follow-up, the structure of the bias is analogous to the DAG illustrated in Fig. 5.2. In this DAG, a predictor (V) of the exposure (X) causes loss to follow-up (L), and a separate predictor (U) of the outcome (Y) also causes loss to follow-up. An analysis that is restricted to those who are not lost to follow-up (L = 0) will suffer from selection bias, as illustrated by the fact that the backdoor pathway X–V–U–Y is unblocked in the DAG (as explained in Chapter 2). Without further analytical adjustments for loss to follow-up (such as analytical adjustment for V), the analysis of the effect of X on Y among L = 0 will be biased.

Such biases are usually difficult to assess in published studies, because the relevant information is often not available or not reported (Example 5.4 and Side Box 5.2). However, reviewers can draw a DAG, such as the DAG illustrated in Fig. 5.2, and determine whether the authors

Fig. 5.2. DAG showing bias due to loss to follow-up in a cohort study: *L*, indicator for loss to follow-up; *U*, unmeasured covariate; *V*, measured covariate; *X*, exposure of interest; *Y*, outcome.



Example 5.4. Bias due to loss to follow-up in an occupational cohort study of flight attendants

Cancer follow-up is frequently based on existing cancer incidence or mortality records. However, national cancer incidence registries are available in only a small number of countries. When cancer incidence is not available through linkage to records, other follow-up methods are needed. An example is a study in the USA of breast cancer among Pan Am flight attendants (Schubauer-Berigan et al., 2015).

Breast cancer incidence in the flight attendant cohort was compared with that in the general population. The incidence cohort included 6093 women who responded to a questionnaire, of whom 134 were proxy respondents in the survey, mostly for deceased cohort members (see Example 5.3).

Information on incident breast cancers was first obtained through self-report of a cancer in the questionnaire. A medical record follow-back of each reported case of cancer was conducted by contacting the physician's office, hospital, or other health-care organization in which the cancer diagnosis was made and obtaining supporting documentation of the diagnosis. Self-reported breast cancers that were refuted by a review of the medical records were not included, but reported cancers that were neither confirmed nor refuted were included. The incidence cohort was also linked to cancer registries in six states, based on the locations of the domiciles for the airline and on common states of residence for the cohort; 82% of the cases of breast cancer in the cohort were verified using medical record follow-back, cancer registry linkage, or both. Loss to follow-up could have occurred if a substantial proportion of the cohort lived in areas without a cancer registry. However, this did not seem to be the situation for this study. (text continues on page 128)

Side Box 5.2. Information that should be reported to enable the assessment of bias due to loss to follow-up

The ideal information that would be reported to enable investigators to determine the presence of bias due to loss to follow-up would be the distribution of exposure, outcome, and confounders, stratified by whether participants were lost to follow-up. Unfortunately, this information will not generally be available, and it will be impossible to know whether the outcome distribution differs by loss to follow-up, because such data are not collected from those who are lost to follow-up. Instead, investigators are limited to examining the distribution of exposures and confounders collected earlier in the study and evaluating whether there are differences in distributions between those who were and were not lost to follow-up. Any differences in loss to follow-up by the exposure or other key variables should be reported and treated as possible sources of selection bias. (text continues on page 128)

adjusted for a sufficient set of variables to reduce selection bias due to loss to follow-up. In the absence of such information, reviewers can conduct a sensitivity analysis for the probable extent and direction of selection bias (due to loss to follow-up) using the methods presented in Section 5.4.4.

5.2.3 Time-zero specification

In the previous section, it is assumed that the source population is followed up for the entire risk period, and that this risk period is properly defined. To explore this concept further, it is necessary to first define the concept of *time zero*. In a randomized controlled trial, this is the time at which a potential study participant meets all of the criteria for inclusion. The inclusion criteria have been applied and treatment has been randomized; at this point, follow-up time (outcome recording) has begun (Hernán et al., 2016). In a cohort study, one should attempt, as much as possible, to align these components, to define a time zero. Time-zero misalignment can sometimes create selection biases (Example 5.5).

5.2.4 Left truncation (prevalent exposures)

Left truncation can result when the effects of exposure occur fairly rapidly after first exposure but study participants are not studied from first exposure. This situation is also known as prevalent exposures, i.e. when follow-up of participants begins after exposure has begun, so cumulative (prevalent) exposure at enrolment is the starting point.

As shown in Example 5.6, hazardous effects have often been missed in cohort studies in which most study participants were only followed up from 10 or more years after first exposure. This type of bias can often lead to paradoxical results, as with HRT use: the people at highest risk die early, leaving the healthiest exposed people to be studied at later time points and suggesting an apparent beneficial effect of the exposure when the study is limited to that group (Flanders and Klein, 2007). The direction and magnitude of the bias are highly context-specific. For example, in a cancer cohort study of an exposure with a long latency period (i.e. the time between exposure and disease induction), there may be little or no bias from left truncation (e.g. if the 5–10 years after first exposure are not included). In contrast, as

Example 5.5. Identifying time zero in an occupational cohort study of flight attendants

In an occupational cohort study, if there is a requirement that eligible participants have worked in the industry for at least 1 month, then time zero will usually be a specified period (1 month) after the start of employment, and follow-up will start from that date. Bias will occur if follow-up time is counted from the start of employment, because person-time will then be counted for the eligibility period, but anyone who dies during that year will be excluded from the study. In the study of the Pan Am flight attendants described in Examples 5.3 and 5.4 (Schubauer-Berigan et al., 2015), participants were eligible for inclusion if they had been employed by the airline for at least 1 year. In the statistical analyses of the association between circadian disruption metrics and breast cancer, follow-up began no earlier than 1 year after the start of employment (other criteria for the start of follow-up were also applied). (text continues above)

Example 5.6. Left truncation as a source of selection bias in studies of hormone replacement therapy

Hernán (2015) has identified left truncation as a source of selection bias in studies of the effects of HRT use in women, where cohort studies and randomized controlled trials initially yielded different findings, with the former showing protective effects and the latter showing increased risks from HRT use. Hernán (2015) showed that this was because the hazardous effects of exposure on cardiovascular disease occurred in the first 5–10 years after first exposure. (text continues above)

shown in Example 5.6, for HRT use, left truncation would produce serious bias because the hazardous effects occurring 5–10 years after first exposure would not be identified.

Key message

In cancer studies, left truncation is of concern mainly when the induction or latency period is likely to be short, for example for most childhood cancers and for some adult cancers, such as leukaemia, or when follow-up begins decades after the start of exposure.

Cohort studies (and corresponding nested case–control studies) that involve left truncation (prevalent exposures) may suffer from selection bias (Danaei et al., 2012). However, Vandenbroucke and Pearce (2015a, b) have argued that although this form of selection bias can occur, the resulting effect is often trivial. This is particularly true for studies of outcomes, such as occupational cancer, where the induction time for the exposure to have an effect can be long (Example 5.7). Moreover, Vandenbroucke and Pearce (2015a, b) have shown that, provided the relevant information is available for each period of time since first exposure, any such left truncation bias can be removed or minimized by stratifying on (and adjusting for) time since first exposure in the analysis (Side Box 5.3).

5.2.5 Insufficient follow-up

The corresponding problem of right truncation occurs when study participants are not followed up for a sufficiently long period after first exposure. For example, if a study involves a risk period of 10 years but is restricted to incident exposures (i.e. participants exposed for the first time during the risk period), then the maximum follow-up time after the first exposure for any study participant will be 10 years, which will be insufficient for most studies of cancer, particularly cancers with long latency, such as many solid tumours. This depends on the research question, but if, for example, the hypothesis is that the exposure can cause cancer 10–25 years after first exposure and

Example 5.7. Left truncation in a population-based cohort study of breast cancer

In a cohort study of night shift work and breast cancer that was based on the Generations Study, middle-aged women were asked at baseline about their exposure to night shift work during the previous 10 years (Jones et al., 2019). The study found no association between being a night shift worker within the previous 10 years and invasive breast cancer (hazard ratio [HR], 0.98; 95% confidence interval [CI], 0.85–1.14). Because of the left truncation in exposure assessment, long-duration night shift workers were included in the exposed group only if they had survived long enough to enter the 10-year recording period and if they were still working night shifts at that time. Furthermore, because night shift work is most common at young ages, the unexposed group could have included an unknown number of women who had worked night shifts at earlier periods in their lives. (text continues above)

Side Box 5.3. Information that should be reported to enable the assessment of bias due to left truncation

The key parameters that should be reported to enable the post-publication assessment of selection bias due to left truncation are the proportions of study participants who were affected by prevalent exposures at baseline and, ideally, for how long these participants had been exposed (minimum, median, and maximum) before follow-up started. Ideally, the reported findings should also be stratified by time since first exposure. For cancer, because of the relatively long induction or latency period, one would expect the exposure–disease association to vary over time since first exposure. In this situation, heterogeneity by time since first exposure is expected and may also account for differences between studies (e.g. if there were different distributions of time since first exposure in different studies). (text continues above)

follow-up has been for only 10 years from first exposure, this represents a selection bias. Thus, Vandenbroucke and Pearce (2015a, b) argue for the inclusion of both incident and prevalent exposures, but with stratification on, and adjustment for, time since first exposure, if appropriate (Example 5.8 and Side Box 5.4).

5.3 Identifying selection bias in case–control studies

The case–control design is a particularly efficient approach for studying rare diseases that can be difficult to study prospectively because a large cohort size, a long follow-up period, or both would be required to accrue enough case participants and attain adequate statistical power. Population-based case–control studies can also be advantageous (Side Box 5.5), because they enable the study of exposures across the whole range of occupations and industries, whereas industry-based cohort studies tend to be focused on a restricted group of agents within a specific setting.

As noted earlier, if a cohort study is based on a particular population over a certain period, selection bias can occur from selection into the study, loss to follow-up, left truncation, or right truncation. All of these biases can occur in a corresponding case–control study based on the same source population followed up over the same period. For example, if the source population for a cohort study is restricted to incident exposures (e.g. the newly employed inception cohort in a particular factory or industry) and the follow-up period is too short, bias due to right truncation can occur. A case–control study based on this source population and risk period will be affected by exactly the same bias.

Additional selection issues can arise in case–control studies, particularly because control participants are selected from the source population and bias may occur as a result of this selection process. Bias may also occur if not all of the case participants in the source population and risk period are selected for recruitment into the study. The focus here is on the inappropriate selection of case or control participants, and on non-participation of case and control subjects. It should be reiterated that it is important to distinguish selection bias from generalizability, as discussed in Section 5.1.

When evaluating the literature with regard to the potential for bias due to the selection of case or control participants, the ultimate focus will often be not only on whether there is bias but also on the potential direction and magnitude of the bias. This chapter first discusses the mechanisms of potential bias, with some examples, before turning to the

Example 5.8. Right truncation in the cohort of atomic bomb survivors in Japan

A classic example to examine the effects of latency and right truncation draws on the studies conducted among survivors of the atomic bombs in Hiroshima and Nagasaki, Japan. Because the radiation occurred at a known time point, this provides a useful example. In an analysis with follow-up from 1950 through 2000, Richardson et al. (2009) showed that there was no evidence of an association between radiation and lymphoma mortality during periods up to 35 years after irradiation. It was only during follow-up periods of 36–45 years and 46–55 years after irradiation that positive associations were observed, pointing to the need for long follow-up to avoid right truncation. (text continues above)

Side Box 5.4. Information that should be reported to enable the assessment of bias due to right truncation

The key parameters that should be reported to enable the post-publication assessment of selection bias due to right truncation are the minimum, median, and maximum lengths of follow-up for the study participants from baseline, as well as the corresponding times since first exposure. As with left truncation (see Side Box 5.3), to enable the assessment of possible bias due to right truncation, the findings should also be stratified by time since first exposure. Once again, heterogeneity by time since first exposure is to be expected for many cancer outcomes. (text continues above)

Side Box 5.5. The population-based case–control study Interphone

Within the four main themes considered in this book to illustrate the concepts of interest (red meat consumption, opium consumption, radiofrequency electromagnetic field (RF-EMF) radiation, and night shift work), examples are often drawn from the Interphone study of RF-EMF radiation exposures (Cardis et al., 2010). This carefully conducted multicentre study included several ancillary and detailed analyses to rule out potential biases. While the study is cited here for illustrative purposes, this should not be considered as a judgement on the quality of the study but, rather, reflects the extensive attention given to methodological issues in the study. Therefore, the study represents a model of careful consideration and discussion of such issues. Most published studies do not report this level of information relevant to selection bias. In this situation, one is usually left with other tools for assessing selection bias, for example through the use of negative control exposures or negative control outcomes (see Section 5.4.2) or hypothetical sensitivity analyses (see Section 5.4.4). (text continues on page 132)

question of direction and magnitude of bias in Section 5.4.4. The initial focus is on relatively simple selection mechanisms that enable the reader to intuit the implied direction of the bias. Section 5.4.4 gives more formal tools to determine the direction and magnitude of bias. Elsewhere in this book, biases are discussed in terms of being towards or away from the null. However, selection bias results in biases that are either upwards or downwards, and in this chapter the result of selection bias is referred to in those terms. For instance, an upward bias (which may result if exposed cases are more likely than unexposed cases to be enrolled in the study) could result in a true odds ratio of 1.5 being estimated as an odds ratio of 2.0, which is both upwards and away from the null. However, the same mechanism could bias a true odds ratio of 0.5 to an estimated odds ratio of 0.8, which is both upwards and towards the null.

5.3.1 Selection of case participants

(a) Source of case ascertainment

Ascertainment of all eligible case participants within a source population can be achieved in several ways, such as using central registry information that is continually updated to include incident cases, or conducting comprehensive active ascertainment of case participants across medical facilities (pathology departments, hospital registries, etc.). Referral by medical sources (treating physicians, clinics, etc.) alone may result in incomplete ascertainment of case participants. To avoid incomplete selection of case participants, information from several sources can be used for cross-validation (Example 5.9).

Depending on the approach being used, cases of more-aggressive or less-aggressive cancers may be missed (Example 5.10). Population-based ascertainment of benign tumours, which are not necessarily included in central tumour registries, can pose a particular challenge. Ascertainment across a very large number of treating institutions may be necessary but is logistically difficult (Example 5.11).

Example 5.9. Cross-validation to improve ascertainment of case participants

To improve the accuracy of case ascertainment of brain tumours in the Interphone study, most study centres used one or more secondary information sources, including medical archives, hospital discharge and billing files, and hospital or regional cancer registries (Cardis et al., 2007). (text continues above)

Example 5.10. Potential bias resulting from differential selection of case participants

In a case–control study of opium consumption and urinary bladder cancer, conducted in the Islamic Republic of Iran (Shakhssalim et al., 2010), an *IARC Monographs* Working Group noted that there appeared to be a selection of case participants with less-aggressive bladder cancer (Table 2.2 in IARC, 2021). Such differential selection of less-severe cases of cancer could introduce bias if, for example, case participants were ascertained from a screening programme in which opium users were less likely to participate. This could occur because of differences in access to health services or in willingness to access them. In such a situation, exposed case participants would be underrepresented in the case–control study, compared with unexposed case participants, and this would bias the observed effect towards the null. (text continues on page 133)

Example 5.11. Potential bias from incomplete case ascertainment of benign tumours

In the Interphone study, many participating centres did not have access to centralized registries of benign parotid gland tumours, and complete case ascertainment would have been problematic (Cardis et al., 2007). As a result, only malignant parotid gland tumours were included in the study. This would not necessarily introduce a selection bias, but it would mean that the findings applied only to malignant tumours and may not be generalizable to benign tumours. (text continues below)

(b) Type of diagnosis confirmation

Studies of cancer usually rely on cases of cancer that have been verified histologically. Rapid access to pathological findings is especially important for cancers with a poor prognosis, and this can preclude the use of central registries. Alternative approaches, which can vary in sensitivity, are sometimes used, depending on the disease and the study setting (Example 5.12).

(c) Exclusion of case participants based on previous history of cancer

In some studies of cancer, patients with previous histories of other cancers are excluded as case participants; this can result in the incomplete inclusion of eligible case participants in the source population and risk period (Example 5.13).

(d) Disease detection issues

Some cancers (e.g. prostate, breast, colon) may be more likely to go undetected in countries where detection is associated with higher SES. The ascertained cases may thus underrepresent subpopulations with lower SES, who in turn may have greater or lesser exposure. An example of this could be lower breast cancer detection among women with lower SES, who may be more often exposed to night shift work. With this selection mechanism, exposed cases would be less likely to enrol than unexposed cases; the observed effect estimate (e.g. the odds ratio) would be biased downwards, and any observed positive effect estimate would be smaller than the true effect estimate due to this selection bias.

(e) Inclusion of prevalent cases of cancer

Cancer case–control studies are usually based on newly diagnosed incident cases (Vandenbroucke and Pearce, 2012). In general, prevalent cases of cancer (i.e. those that were diagnosed at some previous time point) should not be included. However, for some rare tumours with a very prolonged onset, such as chronic lymphocytic leukaemia, it may be difficult to conduct a sufficiently large study without also including prevalent cases. It is sometimes not reported clearly whether a study was

Example 5.12. Potential selection bias arising from different sources of case ascertainment

In most participating countries in the Interphone study, diagnoses were either histologically confirmed or based on unequivocal diagnostic imaging (Cardis et al., 2007). However, in a few countries, only histologically confirmed tumours were included. This could introduce selection bias if a particular exposure were associated with diagnostic imaging. For example, if diagnostic imaging were available only through private hospitals, the case group (identified through histology) might underrepresent cases of cancer in more-affluent patients compared with less-affluent ones, and would also underrepresent exposures associated with affluence, although biasing the odds ratio downwards for these exposures. (text continues on page 134)

Example 5.13. Potential bias from excluding people with previous cancer from the study

In the Interphone study, patients in Denmark who had been found to have had any previous cancer (excluding non-melanocytic skin cancer) were excluded from the study (Cardis et al., 2007). If mobile phone use was associated with other cancers, this exclusion could lead to fewer exposed cases being eligible for the study. If this were the only source of bias, it would bias the observed effect estimate downwards. More probably, such a source of bias would affect the selection of control participants to a lesser extent. Interested reviewers can use the simple methods outlined in Section 5.4.4 to determine the direction of bias. (text continues on page 134)

restricted to incident cases of cancer or also included prevalent cases (Example 5.14).

Selection of case participants in case–control studies can be accomplished by selecting either all eligible cases or a representative sample of those cases. The most common approach is for an investigator to try to enrol all eligible cases (in the source population, over the risk period) in a case–control study. However, it is also possible to conduct a case–control study by selecting a fraction of the eligible cases. In this study design, investigators should sample cases using the same sampling frame used for controls, namely that selection of case participants should be independent of their exposure status. This point is examined in greater detail in Section 5.3.2.

5.3.2 Selection of control participants

(a) Population control participants

Ideally, in case–control studies, case and control participants should represent the same underlying source population over the same risk period. Appropriate selection and recruitment of control participants in a study can be a significant challenge logistically and may pose a threat to study validity. Population-based control participants are usually preferred, and several approaches can be taken to attempt a full population coverage; for example, electoral lists, telephone directories, or lists of general practitioners, where available, could be consulted (any restrictions in availability would apply to the source population and should therefore also be applied to the case participants). Exhaustive recruitment of eligible population control participants is difficult, and response rates in case–control studies of cancer have been shown to decrease over the years. For instance, the median response rate among population control participants in this type of study conducted in 1971–1980 was 75.6%, compared with 53.0% in 2001–2010 (Xu et al., 2018). Although a lower response rate does not necessarily produce selection bias, there is a higher potential for such bias to occur. The most important point to emphasize in selection of control participants is that for an unbiased estimate, control participants should represent the exposure distribution in the source population. A sufficient approach to solving this problem is to sample control participants from the

Example 5.14. Potential bias arising from inclusion of prevalent cases

When reviewing a case–control study of opium consumption and urinary bladder cancer conducted in the Islamic Republic of Iran (Shakhssalim et al., 2010), the *IARC Monographs* Working Group noted that it was unclear whether newly registered cases of cancer might include prevalent cases (IARC, 2021). If prevalent cases were included, this would mean that the overall case group would be weighted towards patients with less-aggressive tumours, because those previously diagnosed with more-aggressive tumours were more likely to have died. If opium consumption caused less-aggressive tumours and these were overrepresented in the study because of the inclusion of prevalent cases, this in turn would produce an increase in the estimated odds ratio compared with that which would have been obtained if the case group had been restricted to patients with only incident tumours. (text continues on page 135)

source population without regard to their exposure status.

Study participants have repeatedly been shown to have a higher education level or higher SES than non-participants (e.g. Fry et al., 2017). Large differences in SES between case and control participants could reflect selection bias if exposure is associated with SES. This may apply particularly to the selection of control participants, for which there are often larger problems of non-response than in the selection of case participants (Example 5.15).

(b) Hospital control participants

In some instances, it is logistically difficult or impossible to enumerate the source population and therefore impossible to recruit control participants at random from the same source population as the case participants. A common alternative strategy is to recruit as the control group patients (in the same source population and risk period) who have other diseases but attend the same health services as the case participants. In recruiting such control participants, it is important to draw on other diseases unrelated to the exposure of interest (so that selection of control participants does not depend on the exposure), because otherwise there is a risk of introducing selection bias. Although it is not possible to remove such selection bias through analysis, the direction of the bias can be predicted based on knowledge of the exposure and relation to the disease in the control group.

Key message

If the exposure of interest is a risk factor for the control disease or the prevalence of the exposure is lower in the source population than among the control participants (for some other reason), then the odds ratio estimate is biased downwards. Conversely, if the exposure of interest is a preventive factor for the control disease or the prevalence of the exposure is higher in the source population than among the control participants (for some other reason), then the odds ratio estimate is biased upwards.

Example 5.15. Indirect evaluation of potential selection bias from differential participation rates in a case–control study

In a case–control study of night shift work and prostate cancer, the sociodemographic characteristics of participants and non-participants, stratified by case–control status, were compared using census-based SES indicators for participants' residential addresses (Barul et al., 2019). The small differences in SES observed between participants and non-participants provided reassurance that there was not major selection bias based on exposure. If such information were not available, one could still attempt to estimate the probable magnitude and direction of any such bias using the quantitative methods outlined in Section 5.4.4. (text continues above)

Control diseases should be selected with caution after the research question and the exposure of interest have been clearly defined. Furthermore, several diseases can be chosen to dilute potential bias introduced by using one particular disease for the control participants, as well as to provide a sufficient sample size to enable sensitivity analyses with various disease control series.

(c) Berkson bias

Berkson bias is a special type of selection bias that may arise when case participants are selected from hospitalized patients and the exposure of interest affects the probability of hospitalization if some case participants are more likely to be hospitalized if they also have another disease (Snoep et al., 2014). Under this scenario, Berkson bias may occur both in studies with population control participants and in studies with hospital control participants. Fortunately, in cancer studies based on incident cases of cancer, the impact of Berkson bias is likely to be reduced, because most (if not all) case participants selected in the hospital will have been hospitalized because of the case disease (Pearce and Richiardi, 2014). Thus, among those case participants, the exposure is not an independent cause of hospitalization; in other words, it is unlikely that a case participant was incidentally discovered among people admitted to the hospital for a different reason. The same logic applies to selection of control participants when control participants are recruited from within a hospital; the control disease should be the cause of hospitalization, rather than being merely present in patients hospitalized for other reasons (see Examples 5.16 and 5.17).

Example 5.16. Evaluating potential Berkson bias in a case–control study

Mohebbi et al. (2021) conducted a case–control study of opium use and head and neck squamous cell carcinoma in 10 provinces in the Islamic Republic of Iran. Included case participants had an incident head and neck cancer and were actively identified through review of admission and treatment information of patients admitted at the cancer care centres of the provinces involved in the study. Control participants were "hospital visitors who were relatives or friends of hospitalized patients in either nononcology wards or who visited the hospital for any reason other than receiving treatment concurrently" (Mohebbi et al., 2021). Berkson bias is unlikely in this study, because all case participants had an incident disease and control participants were not hospitalized. Although Berkson bias may not be a concern in this study, it should be noted that the recruitment of friends as control participants could cause substantial bias, for other reasons. If opium use does, in fact, cause head and neck cancer, we would expect a higher prevalence of opium use among case participants than among the general population. However, friends of hospitalized patients who are opium users may also be more likely to use opium and, as a result, the control series could overestimate the prevalence of opium use in the general population, biasing the observed effect estimate downwards. (text continues above)

Example 5.17. Evaluating potential selection bias from recruitment of hospital-based control participants

The Working Group for *IARC Monographs* Volume 126, on opium consumption (IARC, 2021), evaluated several hospital-based case–control studies, all conducted in the Islamic Republic of Iran. For some of these studies, the Working Group raised concerns about the possibility of selection bias arising as a result of the choice of the control diseases. To avoid this source of selection bias, the disease (or diseases) used to identify hospital-based control participants should be unrelated to the exposure of interest (opium consumption in this example), while it can be affected by other risk factors for the case disease that are unrelated to the exposure of interest. (text continues on page 138)

The direction of Berkson bias can be predicted, theoretically, if the direction of the association between the exposure and the control disease is known or, empirically, if there is information on the prevalence of the exposure in the source population (e.g. the catchment area of the hospital). If the exposure of interest is a risk factor for the control disease or the prevalence of the exposure is lower in the source population than among the control participants (who have the control disease), then the odds ratio estimate is biased downwards; if the exposure of interest is a preventive factor for the control disease or the prevalence of the exposure is higher in the source population than among the control participants, then the odds ratio estimate is biased upwards.

Another source of non-population control participants includes visitors to hospitals (see Example 5.16). In some instances, this is less likely to result in selection bias because the visitor control participants are perhaps more likely than hospitalized control participants to be representative of the general (source) population (Example 5.18); however, great caution should be exercised because hospital visitors could share similar exposure patterns to the case patients being visited.

(d) Using more than one control group

The inclusion of more than one control group allows for a triangulation approach in which the extent and direction of bias is likely to vary across the control groups, and the findings obtained for the different groups can be compared (see Chapter 6). This approach is often used in hospital-based case–control studies in which people with different diseases are recruited to form different control groups (Example 5.19). This topic is discussed further in Section 5.4.4.

5.3.3 Participation of case and control participants

There is a potential for selection bias when both the disease and the exposure status affect participation in the study. This is common in case–control studies, because potential participants typically know their disease and exposure status. In addition, case and control participants may be approached in different settings (e.g. hospitalized case participants and population control participants), and case participants with a poor prognosis might be excluded if they die before recruitment is possible. Furthermore, a person's interest in the study topic may depend on the outcome status (in general, case participants are expected to be more motivated to participate than control participants) as well as on the exposure (some people may believe, for example, that their participation in a study is not essential if they have had no or low exposure); see Example 5.20a.

Example 5.18. Recruiting hospital visitors as control participants

In a case–control study of opium use and oesophageal cancer in the Islamic Republic of Iran, hospital visitors were recruited as control participants (Shakeri et al., 2012). In this study, as noted in Examples 5.22 and 4.14, this control group had an exposure prevalence similar to that observed in the general population of the region, whereas hospital-based control participants in a related study had a higher prevalence of opium use compared with the general population. (text continues above)

Example 5.19. Triangulation across control groups in a study of titanium dioxide exposure

In a study of occupational exposure to titanium dioxide and lung cancer, an analytical control group was recruited that combined a random selection of an equal number of control participants from the general population and from patients with other cancers, to balance the advantages and disadvantages of recruiting population and hospital-based control participants (Boffetta et al., 2001). (text continues above)

Key message

Often, participation does not depend directly on the exposure but is related to factors, such as age, sex, or SES (e.g. if young, working-class men are less likely to participate), that are frequently related to exposure (e.g. occupational exposure to pesticides) (Xu et al., 2018). If those determinants of participation were identified and adjusted for, this selection bias could be controlled as if it were a confounder (from a DAG perspective, this is equivalent to blocking a backdoor pathway that was opened due to conditioning on a collider).

For example, if the outcome was not related to SES in the source population but there was differential participation (between case and control participants) by SES in the study population, then SES would be associated with the outcome in the study population; one can then control for this selection bias, by controlling for SES, just as one would control for confounding (Example 5.20b).

At the time of recruitment, and depending on the method of recruitment, it is sometimes possible to ask people, typically control subjects, who decline to participate a few quick questions about their exposure status in general terms and use this information to identify or model potential bias based on exposure (Example 5.20c).

To mitigate the impact of non-response due to death or severe illness, case–control studies may incorporate proxy interviews with the next of kin of the index participants. Although this approach reduces the potential for selection bias and increases the study power, it may introduce bias through non-differential or differential misclassification (see Section 4.2.3). For this reason, studies involving proxy interviews often include a sensitivity analysis restricted to index interviews (Example 5.20d).

5.4 Tools for assessing and adjusting for selection bias

When a published paper is considered, selection bias can be particularly difficult to assess, because most published studies provide little or no discussion of the potential for selection bias, in contrast to the usually more extensive discussions of the potential for confounding or misclassification. Furthermore, even if the authors of a paper discuss selection bias, the information needed to determine the extent of selection bias (participation rates of cases or controls, with data for exposure or disease status) is generally unavailable. Therefore, one can be left with the impression that selection bias is possible in the study being considered (e.g. because of a low response rate) but have little information to assess whether such bias is likely or its probable magnitude and direction.

Example 5.20a. Potential bias from non-participation in a population-based case–control study

In the Interphone study, a multicentre case–control study of mobile phone use and risk of specific cancer types, the overall participation was 53% for population control subjects, 64% for case subjects with glioma, 78% for case subjects with meningioma, and 82% for case subjects with acoustic neuroma (Cardis et al., 2007; Vrijheid et al., 2009). Of the eligible control subjects identified, 30% refused to participate and 13% could not be traced; the refusal proportion was 11% for all three case participant subtypes, but patients with glioma were more commonly deceased or too ill to participate (15%) than patients with meningioma (2%), patients with acoustic neuroma (0%), or control participants (0%). Because both the proportions of participation and the reasons for non-participation differed between case and control subjects, it is likely that the study was affected by selection bias; however, for this bias to occur, mobile phone use should be associated with participation in case participants, control participants, or both. For example, if people with brain tumours who used mobile phones more often were concerned about the consequences of their phone use and enrolled in the study more often than people with brain tumours who used mobile phones less often, then an upward bias (away from the null if the true OR > 1 and towards the null if the true OR < 1) in the estimated odds ratio would result. (text continues above)

Example 5.20b. Demographic variables as surrogates for examining selection bias

In the Interphone study, the proportions of participation by sex and age group were reported separately for case and control participants (Cardis et al., 2007). In general, these two variables were unrelated to participation, except for a much lower participation among older women with glioma and a slightly higher participation in women than in men among control participants. The study estimates were adjusted for age and sex, which were matching variables. The fact that demographic variables were not related to participation may argue against the presence of selection bias, but this is only indirect evidence, because the exposure, namely the use of mobile phones, might still be a determinant of participation. (text continues on page 139)

Example 5.20c. Use of short questionnaires among non-respondents in a case–control study

Some centres in the Interphone study asked people who declined to participate (30% of the eligible control participants and 11% of the potential case participants) to complete a short non-response questionnaire (NRQ) (Vrijheid et al., 2009). At the 12 centres that asked eligible control subjects to complete an NRQ, 57% (n = 1678) of control group refusers and 2% (n = 26) of other non-participants who were eligible for the control group completed the NRQ. At the nine centres that used the NRQ for potential case participants, 215 potential case participants completed the NRQ, representing 41% of case group refusers and 4% of other non-participants who were eligible for the case group. In both case and control subjects, regular mobile phone use was more common among study participants than among non-participants who completed the NRQ. The differences were large (69% vs 56% among control participants and 66% vs 50% among case participants). The data collected using the NRQ also indicated an association between refusal and lower education level. This variable had already been selected as a potential confounder for inclusion in all multivariable analyses (Cardis et al., 2007). (text continues on page 139)

Example 5.20d. Examining the potential for bias from use of proxy interviews

In the Interphone study, proxy interviews were used for 13% of the case participants (mainly those with gliomas) and 1% of the control participants (Cardis et al., 2007). The exclusion of these participants would have reduced the response proportion among case participants to 59%, not very different from the response proportion observed among control participants. Results of the sensitivity analyses excluding proxy interviews were consistent with the results of the main analyses (INTERPHONE Study Group, 2010). (text continues on page 139)

In this section, tools are discussed that can be used to assess selection bias in published papers when the relevant information is available. Three general types of assessment are considered: (i) substantive knowledge and the use of DAGs; (ii) assessment of selection bias within a single study; and (iii) assessment of selection bias through comparisons across studies. Quantitative sensitivity analysis for selection bias is addressed in Section 5.4.4.

5.4.1 Tool S-1: substantive knowledge and DAGs

Assessing the potential for selection bias requires expert knowledge, usually from previously published studies, and mechanistic knowledge. Ideally, this can be summarized in a DAG (see Chapter 2). For example, in a cohort study, if loss to follow-up is systematically associated with both exposure history and disease status (e.g. as in the healthy worker survivor effect; Checkoway et al., 2004), then there is the potential for selection bias, which can be represented in a DAG, in which conditioning on selection (inclusion in the follow-up) produces an open pathway from exposure to outcome, i.e. collider stratification bias. The DAG will not identify whether such a bias is likely to occur (this depends specifically on the recruitment and retention processes of the particular study) or its probable magnitude and direction (although this can be estimated using signed DAGs; see Section 2.6), but it does provide a framework for considering whether such a bias is possible and evaluating any strategies that the authors may have adopted to minimize, control for, or assess it.

Similarly, in a case–control study, if the response rate is particularly low among control participants, it is possible that selection bias may have occurred if recruitment was related to exposure status (Example 5.21). Again, this bias arises through conditioning on inclusion in the study (it is usually only possible to analyse the data for those who were recruited) and introduces an open pathway from exposure to outcome, i.e. collider bias.

Assessing whether the recruitment of hospital control participants has generated a bias, and, if so, its probable magnitude and direction, requires substantive knowledge from previously published studies, or mechanistic information (Example 5.22).

Example 5.21. Potential selection bias from differential participation in a case–control study

In the Interphone study, almost all exposed groups were found to have lower risks of brain tumours than the unexposed groups. It has been hypothesized (Cardis et al., 2007) that potential control participants who did not own a mobile phone were less likely to participate. If this were the situation, mobile phone use would be overestimated in the control participants, thus producing a downward bias in the estimated odds ratio. (text continues above)

Example 5.22. Potential bias from recruitment of hospital-based control groups

As noted in Example 5.18, Shakeri et al. (2012) conducted a case–control study of opium use and oesophageal cancer in the Islamic Republic of Iran, which involved the recruitment of inpatients in hospitals as control participants. The prevalence of opium use was found to be significantly higher in the hospital control participants than would have been expected on the basis of general population data. One potential explanation for this is that opium use may cause other health problems that result in hospitalization or may be associated with other lifestyle factors that increase the risk of these other health problems. In this situation, the prevalence of opium use in the hospital control participants would be higher than that in the source (general) population, thus producing selection bias. (text continues on page 142)

5.4.2 Tools S-2 to S-6: assessment of selection bias within a study

(a) Tool S-2: negative control exposures

A negative control exposure approach (see also Chapters 2, 3, and 4) involves assessing the association with another exposure that is believed to not be plausibly associated with the outcome under study but is likely to be subject to a similar selection bias (Example 5.23).

(b) Tool S-3: negative control outcomes

A similar approach can be taken with regard to negative control outcomes (Example 5.24). This approach is usually most applicable to cohort studies, because case–control studies are usually based on a single outcome.

(c) Tool S-4: ad hoc reanalysis of published data

In some circumstances, if the necessary information is available, it is possible to reanalyse published results in a manner that potentially reduces selection bias. For instance, if it is thought that there has been selective recruitment with regard to exposure status – for example, if unexposed people are less (or more) likely than exposed people to enrol as control participants – it may still be possible to conduct a dose–response analysis that is restricted to exposed participants (Example 5.25). This relies on the assumption that even if unexposed people were less (or more) likely than exposed people to participate, the level of exposure among those who are exposed does not affect the probability of recruitment. This approach has often been used in occupational epidemiology when risk is compared between people with various levels of exposure rather than between exposed and unexposed people; unexposed people are regarded as an entirely different group (Saracci and Samet, 2010). For example, more valid estimates may be obtained by comparing manual workers across different levels of exposure, rather than by comparing the exposed workers with the general population.

(d) Tool S-5: comparisons with external data

A further approach for assessing selection bias involves making comparisons with external data on the exposure prevalence in the source population (Examples 5.26 and 5.27). This can involve information either on the exposure itself (e.g. pesticide exposure in the general population) or on a surrogate of exposure (e.g. being a farmer).

Example 5.23. Using negative control exposures to examine potential selection bias in a case–control study

In a case–control study of night shift work and breast cancer, any selection pressures (e.g. control participants being less likely to participate if they have never worked night shift) are likely to apply to other non-standard work shifts (e.g. afternoon shift), rather than only to night shift work. If it is well established that afternoon shift work is not associated with breast cancer, then afternoon shift work could serve as a negative control exposure. If a strong association were found between afternoon shift work and breast cancer in the case–control study, this would provide evidence of selection bias, as well as its probable magnitude and direction. (text continues above)

Example 5.24. Using negative control outcomes to examine potential selection bias in a case–control study

If it is well established that the main exposure is not associated with a particular outcome (outcome B) that is different from the main outcome under study (outcome A), then this information can be used to assess selection bias (e.g. due to selective recruitment or loss to follow-up). In particular, if the effect estimate (e.g. odds ratio) is elevated to a similar extent in both the main study outcome and the negative control outcome, this may indicate that the increase in risk for the main study outcome is due to bias. (text continues above)

Example 5.25. Using dose–response analysis to examine potential selection bias in a case–control study

In the Interphone study, almost all exposed groups were found to have lower risks of brain tumours than the unexposed groups. For example, the odds ratio for the lowest exposure group (< 5 hours of cumulative call time) was 0.8, compared with the unexposed group (INTERPHONE Study Group, 2010). It has been hypothesized (Saracci and Pearce, 2008) that potential control participants who did not own a mobile phone were less likely to participate. If this were the situation, mobile phone use would be overestimated in the control participants, thus producing a downward bias (towards the null if the true OR > 1 and away from the null if the true OR < 1) in the estimated odds ratio. One way to investigate this situation is to conduct analyses excluding both case participants and control participants who were not mobile phone users (Cardis et al., 2007). In this study, the odds ratios for meningioma were only slightly changed, whereas those for gliomas became mostly close to (and above) 1 (Saracci and Samet, 2010); the odds ratio for the top decile of cumulative call time increased from 1.40 to 1.82. Saracci and Samet (2010) comment that the direction of these corrections again indicates a contribution of non-participation (selection) bias to the observed low odds ratios. (text continues on page 142)

Example 5.26. Using external data on exposure prevalence to examine potential selection bias in a case–control study of pesticide exposure

In a study of pesticide exposure and soft tissue sarcoma (Smith et al., 1984), control participants who had cancers other than soft tissue sarcoma were recruited, to minimize information bias (because the control participants also had cancer and would have gone through a similar thought process to that of the case participants in terms of the potential causes of their cancer). However, if some of the cancer types in these control participants were also caused by pesticide exposure, selection bias would have occurred due to overrepresentation of pesticide exposure among control participants, thus leading to bias downwards in the estimated odds ratios. Information on pesticide exposure in the general population was not available, but such exposures occur mainly in farming, and information was available on the proportions of workers in various farming groups in the general population. Thus, it was possible to compare the proportions of control participants who were farmers with the expected proportion based on the general population; this comparison showed that it was unlikely that this form of selection bias was occurring (Pearce et al., 1983). (text continues on page 142)

Example 5.27. Using external data on exposure prevalence to examine potential selection bias in a case–control study of opium exposure

As described in Examples 4.14, 5.18, and 5.22, Shakeri et al. (2012) compared the results of two different case–control studies of opium use and oesophageal cancer conducted in the same region by a single research group. In one study, hospital-based control participants were recruited, whereas the other study involved control participants drawn from the neighbourhood. The prevalence of opium use was also estimated from a cohort that was enrolled in the same geographical area and therefore probably represented the source population for the study. The standardized opium consumption prevalence was 0.17 in the cohort, 0.16 in the neighbourhood control participants, and 0.23 in the hospital-based control participants, suggesting that the neighbourhood control participants were more representative of the study base population for this exposure. (text continues on page 142)

(e) Tool S-6: using several control groups

It is unusual for studies to involve more than one comparison or control group, but when this is done the information obtained can be used to assess the potential for selection bias. This applies particularly when the various control groups are expected to produce biases in opposite directions, as in Example 5.28.

5.4.3 Tool S-7: assessment of selection bias through comparisons across studies

Selection bias can also be assessed by making comparisons across studies (Example 5.29). This applies particularly when similar studies have been conducted in the same population (e.g. cohort studies involving the same industry or the same group of workers, or case–control studies conducted in the same populations). However, comparisons can also be made between studies conducted in different populations where it is reasonable to assume that the strength of the main exposure–outcome association is likely to be similar. For example, one might compare the findings from studies in which control participants were recruited from the general population with those from studies in which control participants with diseases other than the disease

Example 5.28. Using triangulation of findings from different control groups to examine biases in a case–control study of pesticide exposure

In a study conducted in New Zealand to investigate a possible association between phenoxy herbicides and non-Hodgkin lymphoma (Pearce et al., 1986), control participants were recruited from the general population and also from among people who had other cancers. The assumption is that if there were any recall bias, this would be more likely in the general population control participants (who may not recall all of their exposures), and the comparison with this control group would produce artificially high odds ratios (i.e. bias upwards). Conversely, the recruitment of control participants who had other cancers would be expected to minimize recall bias, but there might be selection bias and hence a bias downwards in the estimated odds ratio (see previously) if some of the cancer types in these control participants were also caused by phenoxy herbicides. A key issue is that these biases would operate in different directions, allowing the possibility of triangulation of the findings with the two control groups. In fact, the study produced similar results for each control group, indicating that both recall bias and selection bias were unlikely to be important problems in this study. (text continues above)

Example 5.29. Comparisons across studies to examine potential biases in case–control studies

As described in Examples 4.14, 5.18, 5.22, and 5.27, Shakeri et al. (2012) compared the results of two different case–control studies of opium use and oesophageal cancer conducted in the same region by a single research group. Case definition and enrolment of case participants were the same in the two studies. However, the selection of control participants differed: in one study, hospital-based control participants were recruited, whereas the other study involved control participants drawn from the neighbourhood. The prevalence of opium use was found to be significantly different between the hospital and neighbourhood control participants, but the prevalence of tobacco use did not differ between these groups. Consequently, the inference drawn for the association between oesophageal cancer and tobacco use did not differ between the studies, but that for opium use did (IARC, 2021). In the study with neighbourhood control participants, opium use was associated with a significantly increased risk of oesophageal cancer (adjusted OR, 1.8; 95% CI, 1.2–2.7), while in the study with hospital control participants, this was not so (OR, 1.1; 95% CI, 0.6–1.9). This indicates that selection bias is likely to have occurred, and to have been substantial, in the study with hospital control participants, although the possibility that neighbourhood control participants may be prone to other selection factors cannot be ruled out. (text continues above)

under investigation were enrolled. Such comparisons across studies (triangulation) are discussed further in Chapter 6.

5.4.4 Tool S-8: selection bias adjustment

In this section, an approach to conducting sensitivity analyses for selection bias is demonstrated, with a detailed worked example using methods described in more detail in Fox et al. (2021), beginning with Example 5.30a.

Selection bias is, mathematically, the easiest bias to adjust for. Table 5.1 illustrates a common way in which selection bias occurs in case–control studies. If A = 100 people are eligible for recruitment to the exposed case group in the population but only s_{11} = 70% of them participate in the study, we would have a = 70 participants. The bias parameter s_{11} is the selection probability for exposed case participants. There are three other selection probabilities – for the unexposed case participants, exposed control participants, and unexposed control participants – that determine which data are observed in a study. These types of parameters, which dictate the extent of the bias in the data, are referred to as bias parameters (Side Box 5.6).

If these four selection probabilities are known, it is easy to divide the observed cell counts by the selection probabilities to recover the 2 × 2 table that would have been observed in the absence of selection bias (assuming that the correct selection probabilities are specified): $A = a/s_{11}$, $B = b/s_{10}$, $C = c/s_{01}$, and $D = d/s_{00}$. Unfortunately, these bias parameters are generally unknown, because they require information on the exposure prevalence among case and control participants in the general population – information that, if it were available, would generally obviate the need to conduct a bias adjustment in the first place. In some situations, selection probabilities may be available from ancillary studies, but these situations are limited. When precise information on selection probabilities is lacking, it is common to choose a range of plausible values for each of the four parameter values and conduct a bias analysis over the combination of values. This is referred to as a multidimensional bias analysis (introduced in Section 4.3.2).

Often, study publications give an overall response or participation rate for case and control participants, and this can be used to reduce the number

Example 5.30a. Identifying potential selection bias in a nested case–control study of breast cancer

The Working Group for *IARC Monographs* Volume 124, on night shift work and cancer (IARC, 2020), noted a potential for selection bias in the findings of O'Leary et al. (2006), who had conducted a case–control study of shift work and breast cancer as part of the larger Long Island Breast Cancer Study Project. Case participants were residents of Long Island, New York, who had received diagnoses of incident occurrences of breast cancer between 1 August 1996 and 31 July 1997. Control participants were age-matched to case participants. Control participants younger than 65 years were recruited through random-digit dialling, and those aged 65 years or older were selected from Medicare enrolment lists. Both case and control participants were restricted to people who had lived at the same residence for 15 years or longer. O'Leary et al. (2006) reported that any overnight shift work was inversely associated with breast cancer (OR, 0.55; 95% CI, 0.32–0.94). These results are implausible, based on other reported findings, and it is therefore useful to consider whether the observed protective effect could in part be due to selection bias. The original Long Island Breast Cancer Study Project, within which this study was nested, reported response rates of 82.1% for case participants and 62.8% for control participants (Gammon et al., 2002). O'Leary et al. (2006) reported participation rates for their substudy of 87% for case participants and 83% for control participants. The overall participation rates in the shift work study were unavailable, because the original Long Island Breast Cancer Study Project did not limit enrolment to people who had lived at the same residence for at least 15 years, whereas the substudy on shift work did. Nonetheless, overall rates can be approximated by multiplying the two sets of response rates, yielding an overall response rate of 71.4% for case participants and 52.0% for control participants. Thus, there is certainly a potential for selection bias. (text continues above)

Table 5.1. True and observed cell counts in a case–control study with selection bias[a]

	True cell counts		Observed cell counts	
	Exposed	**Unexposed**	**Exposed**	**Unexposed**
Case participants	*A*	*B*	$a = A \times s_{11}$	$b = B \times s_{10}$
Control participants	*C*	*D*	$c = C \times s_{01}$	$d = D \times s_{00}$

[a] Uppercase letters, unobserved true cell counts; lowercase letters, observed cell counts; s_{ce}, selection probability by case status (c = 0, 1) and exposure (e = 0, 1).

Side Box 5.6. Information that should be reported to enable the assessment of selection bias using sensitivity analysis

The key parameters that should be reported to enable the post-publication assessment of selection bias using sensitivity analysis are the bias parameters shown in Table 5.1, i.e. the selection probabilities for exposed case participants, unexposed case participants, exposed control participants, and unexposed control participants. In some studies, it may be possible to report this information, or proxies for it, if it is available for the source population, and the distribution of these factors (case or control status; exposed or unexposed status) in the study population and the source population can be compared. However, this is rarely the situation; typically, the best that can be done is to hypothesize the probable values (or a range of values) for the four bias parameters shown in Table 5.1 and then conduct the sensitivity analyses covered in this section. (text continues on page 145)

of bias parameters that need to be specified in a sensitivity analysis. For instance, if the overall response rate among case participants is s_{case} and a value for the participation rate among exposed case participants, s_{11}, is specified, then the participation rate among unexposed case participants, s_{10}, can be calculated as

$$s_{10} = \frac{b}{\frac{a+b}{s_{\text{case}}} - \frac{a}{s_{11}}} \qquad (5.1)$$

A similar equation exists for the control participants, if the overall response rate among control participants, $s_{control}$, is known:

$$s_{00} = \frac{d}{\frac{c+d}{s_{\text{control}}} - \frac{c}{s_{01}}} \qquad (5.2)$$

These two equations can be used to implement a sensitivity analysis for selection bias that only requires plausible values to be specified for two remaining unknown parameters: the selection probability among exposed case participants (s_{11}) and the selection probability among exposed control participants (s_{01}), as in Example 5.30b.

These methods enable the researcher to judge how much the point estimate can change after adjusting for selection bias, assuming that the bias parameters are correctly specified, but they do not incorporate uncertainty due to random error. Fortunately, relatively simple procedures can be used to produce interval estimates around the bias-adjusted effect estimates; indeed, the typical variance estimates (e.g. using the delta method) for the log odds ratio that would be calculated from the biased data can be used directly (Example 5.30c).

Quantitative bias analysis methods to adjust effect estimates for selection bias are easily implemented, but the user should be cautious, for several reasons. The first is that the methods rely on accurate specification of the bias parameters. Incorrect guesses of the selection probabilities will result in incorrect bias adjustments. Furthermore, although it may be tempting to assume that if the specified selection probability is close to the truth then the bias-adjusted result will be close to unbiased, this turns out not to be true in general. Having a bias parameter that is close to the true selection probability may still result in a badly biased adjusted effect estimate. The best solution to this problem is to conduct a multidimensional bias analysis (such as that in Table 5.2) and determine the sensitivity of the adjusted effect estimate to changes in the bias parameters.

Example 5.30b. Quantitative bias analysis to examine potential selection bias in a nested case–control study of breast cancer

The estimated response rate among case participants (s_{case} = 0.714) and among control participants ($s_{control}$ = 0.520) can be used to implement a quantitative bias analysis for selection bias in the shift work study of O'Leary et al. (2006). To begin the quantitative bias analysis, the crude 2 × 2 data are abstracted from the paper (a = 26, b = 313, c = 50, and d = 321). As reported in O'Leary et al. (2006), control participants who reported a history of overnight shift work were younger, had a lower household income, and were less likely to have had a mammogram than control participants who had never engaged in overnight shift work. It is assumed that both eligible case participants and potential control participants who had engaged in overnight shift work were less likely to participate in the study than women who had not engaged in overnight shift work. That is, it is assumed that $s_{11} \leq s_{10}$ and $s_{01} \leq s_{00}$. Furthermore, it is assumed that women with incident breast cancer are at least as likely to participate in the study as those without breast cancer: $s_{11} \geq s_{01}$ and $s_{10} \geq s_{00}$. To conduct a sensitivity analysis, we choose a range of values of the bias parameters s_{11} and s_{01} compatible with these assumptions. For example, the selection probability among exposed case participants is specified as slightly lower than the overall response rate among case participants, s_{11} = 0.7. Similarly, the response rate among exposed control participants is specified as slightly lower than the overall response rate among control participants, s_{01} = 0.5. With these values, the selection probability among unexposed case participants can be calculated as

$$s_{10} = \frac{313}{\frac{26+313}{0.714} - \frac{26}{0.7}} = 0.715 \qquad \text{(E5.1)}$$

Similarly, the selection probability among unexposed control participants can be calculated as

$$s_{00} = \frac{321}{\frac{50+321}{0.520} - \frac{50}{0.5}} = 0.523 \qquad \text{(E5.2)}$$

With the four selection probabilities specified, a selection-bias-adjusted odds ratio can be calculated:

$$\mathrm{OR}_{\mathrm{adj}} = \frac{\frac{a}{s_{11}} \times \frac{d}{s_{00}}}{\frac{b}{s_{10}} \times \frac{c}{s_{01}}} = \frac{\frac{26}{0.7} \times \frac{321}{0.523}}{\frac{313}{0.715} \times \frac{50}{0.5}} = 0.52 \qquad \text{(E5.3)}$$

For this set of bias parameter values, one would expect that in the absence of selection bias, approximately the same protective effect of overnight shift work would have been observed (OR_{adj} = 0.52 vs OR_{crude} = 0.53). Note that the second odds ratio is calculated directly from the cells of the observed 2 × 2 table and does not adjust for any confounders. This result is the first row of Table 5.2. This calculation is repeated for s_{11} = {0.7, 0.6, 0.5, 0.4} and s_{01} = {0.5, 0.4, 0.3}. The spreadsheet used in this example is provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat). No combination of these selection probabilities leads to a bias-adjusted odds ratio that supports a harmful effect of overnight shift work. Most bias parameter combinations lead to more protective bias-adjusted effects; only bias parameters that may be viewed as less plausible, such as those with higher participation rates among exposed control participants than among exposed case participants, lead to adjusted effects near the null. These results suggest that for these bias parameter values, selection bias is not likely to be responsible for the observed protective effect of shift work.

Example 5.30b. Quantitative bias analysis to examine potential selection bias in a nested case–control study of breast cancer (continued)

Table 5.2. Sensitivity analysis for overnight shift work and incident breast cancer from a case–control study[a]

Bias parameters				Bias-adjusted OR[b]	95% confidence interval
s_{11}	s_{10}[b]	s_{01}	s_{00}[b]		
0.7	0.715	0.5	0.523	0.52	(0.32–0.86)
0.7	0.715	0.4	0.545	0.40	(0.24–0.66)
0.7	0.715	0.3	0.587	0.28	(0.17–0.46)
0.6	0.725	0.5	0.523	0.62	(0.37–1.01)
0.6	0.725	0.4	0.545	0.47	(0.29–0.78)
0.6	0.725	0.3	0.587	0.33	(0.20–0.54)
0.5	0.740	0.5	0.523	0.75	(0.46–1.24)
0.5	0.740	0.4	0.545	0.58	(0.35–0.95)
0.5	0.740	0.3	0.587	0.40	(0.25–0.66)
0.4	0.764	0.5	0.523	0.97	(0.59–1.60)
0.4	0.764	0.4	0.545	0.75	(0.45–1.23)
0.4	0.764	0.3	0.587	0.52	(0.32–0.86)

OR, odds ratio.
[a] Using an overall response rate among case participants of $s_{case} = 0.714$ and an overall response rate among control participants of $s_{control} = 0.520$ and the observed cell counts ($a = 26$, $b = 313$, $c = 50$, and $d = 321$).
[b] s_{10} and s_{00} and the adjusted odds ratio are calculated using the formulae given in this section, conditional on the observed cell counts and overall response rates.
Source: O'Leary et al. (2006).

(text continues on page 146)

Example 5.30c. Confidence interval estimation when quantifying selection bias for a nested case–control study of breast cancer

In the study by O'Leary et al. (2006), the variance estimated from the crude data is

$$\mathrm{Var}(\log \mathrm{OR}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} = \frac{1}{26} + \frac{1}{313} + \frac{1}{50} + \frac{1}{321} = 0.06 \quad \text{(E5.4)}$$

This variance can be used in conjunction with the bias-adjusted effect estimates derived in Example 5.30b. For example, in the first row of Table 5.2, the bias-adjusted odds ratio is 0.52, and the 95% confidence interval can be calculated as

$$\ln(0.52) \pm 1.96 \times \sqrt{0.065} = (0.32, 0.86) \quad \text{(E5.5)}$$

Similar calculations can be included for each row of Table 5.2 to generate bias-adjusted interval estimates. (text continues on page 146)

A second cause for concern is that the methods presented here admit no uncertainty about the bias parameters; they assume complete confidence in the parameter value. It is possible to specify a distribution for each bias parameter, with that distribution representing the investigator's uncertainty regarding the value of the bias parameter, and then to conduct a probabilistic bias analysis (Example 5.30d).

In a probabilistic bias analysis, uncertainty is incorporated into the bias parameter by repeatedly sampling selection probabilities from each of the four bias parameter distributions. Each set of sampled bias parameters is used to bias-adjust the observed table, as before. Finally, to incorporate the conventional random error, the variance should be based on the non-bias-adjusted cell counts, as calculated previously. This approach is iterated a large number of times, and the resulting estimates are summarized by an overall bias-adjusted estimate (the median of the bias-adjusted results) and an uncertainty interval (the 2.5th and 97.5th percentiles of the bias-adjusted results). This approach can easily be implemented in Excel, R, Stata, or SAS. To find the bias analysis estimate given in Example 5.30e, Excel spreadsheets were used, with 1000 iterations (https://sites.google.com/site/biasanalysis/Home; Fox et al., 2021); the spreadsheet is provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat).

5.5 Other miscellaneous biases

In this final section, several biases are considered that do not necessarily fit neatly into the categorization of biases comprising selection bias, information bias, and confounding.

Example 5.30d. Probabilistic bias analysis to examine potential selection bias in a nested case–control study of breast cancer

In the study by O'Leary et al. (2006), one might believe that the selection probability among the exposed case participants is between 0.6 and 0.8, with 0.7 the most likely selection probability; one could then parameterize this belief as a triangular distribution with a minimum of 0.6, a maximum of 0.8, and a mode of 0.7. This distribution should capture a well-informed belief about the distribution of plausible selection probabilities among the exposed case participants. Similarly, a distribution for each of the three other selection probabilities could be parameterized. For the purposes of this example, the selection probabilities from the first row of Table 5.2 are used. It is assumed that the mode of each distribution is the selection probability given in the table (s_{11} = 0.7, s_{10} = 0.715, s_{01} = 0.5, and s_{00} = 0.523). For simplicity, it is assumed that the minimum of each of the distributions is 0.1 below the mode and the maximum is 0.1 above the mode (e.g. the distribution for s_{01} is centred at 0.5 and has a minimum of 0.4 and a maximum of 0.6). (text continues above)

Example 5.30e. Applying probabilistic bias analysis results to estimated odds ratios in a nested case–control study of breast cancer

For the study by O'Leary et al. (2006), the probabilistic bias analysis returns an odds ratio of 0.52 (95% credibility interval, 0.29–0.91). The point estimate is identical to the point estimate obtained from the simple bias analysis; this will generally be the situation whenever the bias parameter distribution is symmetrical around the mode. The interval estimate for the probabilistic bias analysis is larger than that for the simple quantitative bias analysis; this will generally be the situation, because the intervals for the former analysis incorporate additional uncertainty around the bias parameters. (text continues above)

5.5.1 Healthy worker biases

There are two types of healthy worker bias. The first type is healthy worker hire bias, which occurs when relatively healthy individuals in an occupational population are compared with the general population; it may lead to downward bias in relative mortality measures (e.g. for all causes or for all cancers) (Checkoway et al., 1989, 2004). The second type is healthy worker survivor bias, which occurs because workers who are healthy are more likely to stay employed for longer, thus experiencing the greatest amount of exposure (Pearce et al., 1986). Because of these two selection processes, an occupational population is usually inherently non-comparable with the general population with which it is typically compared in occupational cohort studies. This occurs even if participants continue to be followed up after they leave employment, because they are likely to have lower lifetime cumulative exposure than those who remain in employment. Although healthy worker bias is most commonly discussed in terms of occupational cohort studies, the same issues of bias apply to other study designs (such as nested case–control and cross-sectional studies) that are based on the experience of a cohort over time.

Some authors regard healthy worker bias as an example of selection bias, because of the selection of an inappropriate comparison population (i.e. comparing the general population with a healthy employed population) or conditioning on employment in the industry. Others regard it as an example of confounding, because employed people and those who remain in employment are generally healthier than the rest of the source or general population with which they are being compared (Checkoway et al., 2004; Keil et al., 2015). In the context of this book, healthy worker bias can be regarded as confounding, because it arises from inherent differences between employed and non-employed subgroups in the source population. Therefore, it is also addressed in Chapter 3.

5.5.2 Immortal time bias

Immortal time bias arises if the definition of one of the two exposure groups that are compared within a study is specified incorrectly, such that there is a period during which members of that exposure group accumulate person-time but will not be included in the study if they experience the outcome (Hanley and Foster, 2014). A good example of this was presented as far back as the 1840s by William Farr: generals and bishops live longer than curates and soldiers, but only because one has to reach a certain age to hold such a position (Farr and Humphreys, 1885). This can be regarded as a type of selection bias (related to time-zero specification, described in Section 5.2.3), because some study participants are only included in the analysis if they survive up to a certain time point, but if they do, their person-time up to that point is incorrectly included in the data analysis. Although this issue may seem obvious, this error seems to reappear in epidemiology, and immortal time bias has led to seriously flawed results (Example 5.31).

5.5.3 Reverse causation and protopathic bias

Reverse causation occurs when the exposure changes after the disease of interest occurs or is caused by the diagnosis of the disease. This can be viewed as a type of differential information bias, because exposure has been measured at the wrong time (i.e. too close to the occurrence of disease) and is therefore misclassified. The easiest way to avoid reverse causation is to use a prospective cohort study design, in which a condition of enrolment in a study is not having cancer, perhaps after an initial period to allow for the appearance of cancers that were latent but not yet diagnosed, and then to assess exposure. In case–control studies, reverse causation may occur when there is not careful assessment of the timing of exposure and confirmation that the disease occurs after the occurrence of exposure. One method of evaluating the effect of reverse causation is to exclude individuals who only recently experienced the exposure of interest (Example 5.32).

Protopathic bias is related to reverse causation and is often included in the definition of reverse causation. However, it differs in that the occurrence of disease does not directly affect exposure status. Rather, protopathic bias occurs indirectly when a symptom of the undiagnosed disease causes a change in the exposure of interest in the case participants. Protopathic bias can occur in both cohort and case–control studies (Example 5.33).

In cohort studies of cancer types for which survival is poor, the exclusion of patients who were diagnosed within the early period of follow-up can provide evidence about the extent of

Example 5.31. Immortal time bias in a registry study related to solar radiation exposure

Immortal time bias was observed in a registry study of skin cancer in Denmark (Brøndum-Jacobsen et al., 2013; Lange and Keiding, 2014). The researchers aimed to investigate any beneficial effects of sun exposure on longevity, but because they did not have access to information on sun exposure, they chose people with a diagnosis of skin cancer as a proxy for high sun exposure. The comparison group was all people in Denmark without a diagnosis of skin cancer, and follow-up started at age 40 years. Whereas people in the comparison group were at risk of dying from this age onward, it was impossible for people in the skin cancer group to die before the age of diagnosis, which was, on average, 68 years. The immortal time bias led to people with skin cancer having half the mortality risk of people without skin cancer (relative risk, 0.52), and the study received great attention in the media in Denmark, with front pages stating that sunbathers live longer. In such a study, the correct analysis would be to allow people to change exposure status as they proceed through the study period (this is equivalent to using a time-dependent variable in a Cox model; Pearce et al., 1988). Thus, in this situation, the people with skin cancer should have been considered as part of the unexposed group until they received a diagnosis, and the results of the analysis would have been very different. (text continues on page 150)

Example 5.32. Examining reverse causation in a case–control study of oesophageal cancer

In a case–control study of oesophageal cancer and opium use, there was concern that reverse causation may partially explain the odds ratio of 2.00 (95% CI, 1.39–2.88), if people who developed cancer had a subsequent increased likelihood of taking up opium use. Therefore, Nasrollahzadeh et al. (2008) restricted the analysis to users who had reported use earlier than 1 year before cancer diagnosis; this gave an odds ratio of 1.92 (95% CI, 1.30–2.84), indicating that reverse causation is unlikely to explain the association. (text continues on page 150)

Example 5.33. Examining protopathic bias in case–control studies of opium use and cancer

Opium consumption is an excellent example of an exposure that may be affected by protopathic bias in studies of cancer. In this case, the symptoms of undiagnosed cancer may motivate the patient to self-medicate with opium, making it appear that opium use increases the risk of disease. In studies of opium use and lung cancer, one of the causes of protopathic bias is related to the antitussive properties of opium. Because one of the early symptoms of lung cancer is coughing, the use of opium to ameliorate these symptoms may introduce protopathic bias. In this situation, because tobacco smoking is related to both coughing and lung cancer, controlling for smoking will minimize the risk of protopathic bias. (text continues on page 150)

protopathic bias (Example 5.34). The impact of protopathic bias is more difficult to assess for cancer types for which survival times are longer.

5.5.4 Inappropriate control for a collider (other than selection into the study) in the analysis

Bias can also arise from inappropriate control for a collider (other than selection into the study) (Pearce and Lawlor, 2016), even if 100% of the source population has been recruited into the study (and therefore there cannot be selection bias). Briefly, controlling for any collider can open a backdoor pathway involving that collider, and the resulting bias can only be controlled by controlling for at least one other variable on the same backdoor pathway (Example 5.35).

5.5.5 Biases in biomarker exposure measures

Biomarkers are now extensively used in cancer epidemiology. Within the concept of the exposome, their application has widened to incorporate new high-throughput techniques to evaluate exposure or intermediate pathways and preclinical disease markers (e.g. Wild, 2005). In the context of this book, we consider mostly biomarkers of exposure, i.e. measurements in body fluids or other tissues that correlate with an environmental exposure or an exposure mixture. Biases arising from the use of biomarkers can most commonly be regarded as information bias, but these issues are considered here because they also relate to reverse causation. In contrast, the appropriate use of biomarkers can help to avoid or minimize information bias.

Biomarkers can be used as direct measures of exposure in study participants and are frequently used in a subpopulation to develop exposure models that are then applied to the whole study population by modelling using proxies of exposure (Example 5.36).

Like for any other exposure measured through questionnaires or other methods, errors in biomarker measurements can result from both non-differential and differential misclassification (Fig. 5.4).

(a) Non-differential errors in biomarker measurements

In a case–control study of breast cancer (Mukherjee Das et al., 2022), using urinary concentrations of short-lived chemicals (e.g. phthalates) would introduce non-differential misclassification because of extreme time-related misclassification. The time window of interest for a chronic disease, such as breast cancer, could be 10–20 years before clinical disease diagnosis, while the biomarker would measure exposure only during the previous few weeks. It is unlikely that breast cancer status would affect the performance of the biomarker test or alter levels of phthalates; thus, this is a non-differential misclassification mechanism. Chapter 4 describes tools to assess the direction and magnitude of non-differential biases in continuous exposures.

Example 5.34. Evaluating the potential impact of protopathic bias in a study of pancreatic cancer

In a cohort study of prognostic factors for pancreatic cancer, for which survival is poor, Sheikh et al. (2020) evaluated the potential impact of protopathic bias by excluding any participant who had started using opium in the 2 years before receiving a diagnosis. They found minimal impact on the results when the few participants who had started using opium recently before diagnosis were excluded. (text continues above)

Example 5.35. Inappropriate adjustment for a collider

Richiardi et al. (2008) provide an example of inappropriate adjustment for SES in occupational cancer studies. They consider the scenario where SES is not a cause of the cancer under study but is associated with other occupational factors (apart from the main exposure) that are causes of the cancer under study. In this situation, adjustment for SES can open a backdoor pathway involving the other occupational factors, and thus bias the effect estimate for the main occupational exposure under study. (text continues above)

Example 5.36. Modelling using biomarker-based proxies of exposure in a study of herbicide exposure

In the IARC cohort of phenoxy herbicide (Agent Orange) workers who were exposed to dioxins that are contaminants of the herbicides (Saracci et al., 1991; Kogevinas et al., 1997), several studies were conducted among industrial workers and professional sprayers to measure the most toxic dioxin compound, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), in blood samples. In most workers, measurements were made several years after the end of employment (IARC, 1997). Measurement after the end of employment can be problematic for most chemical exposures, because of the short half-life of most compounds; the chemicals are eliminated from the body during a relatively short time (hours, days, or a few months). Dioxins, like other persistent organic compounds and some metals and radionuclides, have a long half-life, frequently longer than 5 years. In the dioxin cohorts, levels of TCDD since first exposure could be reconstructed by modelling, using information from individual job records and individual measurements of blood levels of TCDD (Fig. 5.3). The studies in subsamples indicated a strong correlation of TCDD levels with duration of employment in jobs or industries with potential exposure to TCDD; it was also observed that TCDD levels increased only after substantial exposure to the herbicides, approximately after at least 1 year of exposure. Exposure models were then developed for all the cohort participants, based essentially on information on duration of exposure. (text continues on page 152)

Fig. 5.3. Serum levels of TCDD, adjusted for lipids, in 253 workers in the USA, as a function of years of exposure

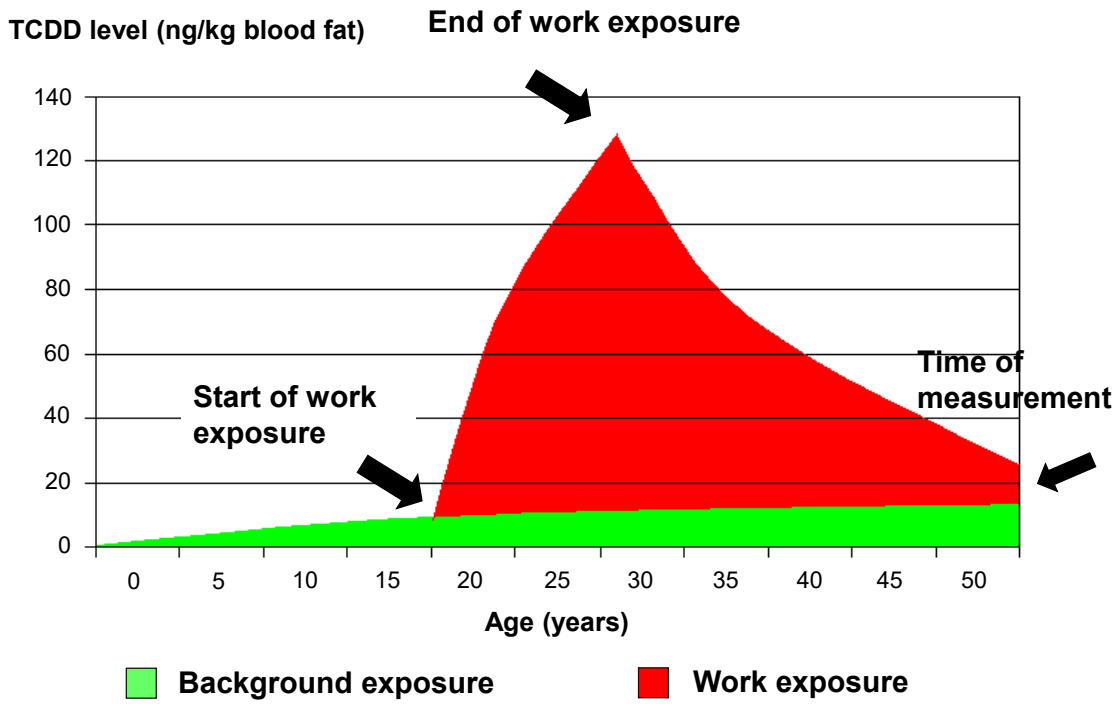
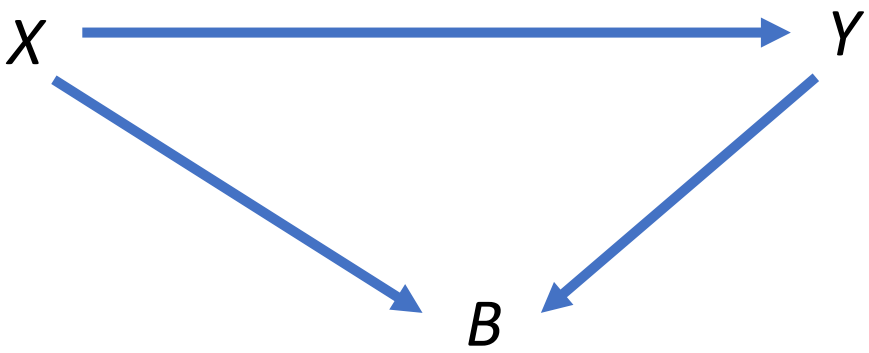


Fig. 5.4. If *X* (exposure) is associated with *B* (biomarker) and there is measurement error of *B*, this would induce non-differential misclassification. If the disease (*Y*) affects the levels of *B*, this would induce differential misclassification.



(b) Differential errors in biomarker measurements

If the biomarker levels were affected by the disease, this could also introduce differential misclassification in a case–control study, because levels among case participants would depend on disease status. This has been described in relation to measurements of chemicals in cancer types with poor prognosis, for example where disease could have affected weight and consequent mobilization of fat tissue, where several persistent compounds are stored in the body. Similarly, the possibility of differential misclassification has been raised in relation to tumours affecting immune status, for example measurements of infectious agents through antibodies, the production of which could be affected by the disease (Aguilar et al., 2017), as in Example 5.37. Section 4.2.3 describes tools for assessing the direction and magnitude of bias from differential exposure.

5.6 Summary

In summary, selection bias can occur because of differences between the study population and the source population. Selection bias can arise through various mechanisms, such as incomplete recruitment from the source population or loss to follow-up. This selection bias is distinct from issues of representativeness or generalizability or transportability, which relate to comparisons between the target population and the source population.

In general, selection bias occurs as a result of incomplete recruitment, if selection depends differentially on exposure and disease status (e.g. if exposed case participants are more or less likely than other groups to be recruited) and if this incomplete

Example 5.37. Differential errors resulting from use of biomarkers in studies of Burkitt lymphoma

Infection with Epstein–Barr virus is a primary cause of endemic Burkitt lymphoma, a common neoplasm in children in Africa. An ecological association has been reported between endemic Burkitt lymphoma and the prevalence of malaria due to infection with *Plasmodium falciparum* (IARC, 2013). In a case–control study of Burkitt lymphoma in children in Malawi, blood levels of antibodies to both Epstein–Barr virus and *P. falciparum* were evaluated, and it was found that there was a strong association with Epstein–Barr virus, a moderate association with *P. falciparum*, and an additive interaction of both infections. However, the observed associations with the two infections could be due to differential misclassification, because antibody levels could be different for children with and without Burkitt lymphoma, particularly if reverse causation was involved, i.e. if having Burkitt lymphoma increased the risk of being infected with malaria. (text continues above)

recruitment is not adjusted for in the analysis. In cohort studies, important mechanisms for selection bias include non-response at baseline, loss to follow-up, left truncation, right truncation, and immortal time bias. In case–control studies, all of these biases are possible; in addition, bias could occur through inappropriate selection of control participants (e.g. a control group that does not provide a valid estimate of the exposure history in the source population).

Qualitative tools for assessing the existence, direction, and magnitude of selection bias include the use of negative control exposures, negative control outcomes, ad hoc reanalyses of published data, comparisons with external data, and the use of several control groups. All of these can be regarded as types of triangulation. Quantitative methods also exist for sensitivity analyses that involve adjusting for hypothesized selection bias. Although these calculations are relatively easy to implement, it is often the situation that there will not be adequate information to specify bias parameters for a range of possible selection effects.

Thus, as noted in Chapter 1, one of the primary questions posed to *IARC Monographs* Working Group experts is, "Can we reasonably rule out selection bias as an explanation for an observed exposure–cancer association?" This can be particularly difficult to assess, because most published studies provide little or no discussion of the potential for selection bias, in contrast to the usually more extensive discussions of the potential for confounding (Chapter 3) or misclassification (Chapter 4). Therefore, it is important that authors, and editors, are encouraged to report the information that is required for a valid assessment of the potential, direction, and magnitude of possible selection bias, as described in Chapter 7.

References

Aguilar R, Casabonne D, O'Callaghan-Gordo C, Vidal M, Campo JJ, Mutalima N, et al. (2017). Assessment of the combined effect of Epstein–Barr virus and *Plasmodium falciparum* infections on endemic Burkitt lymphoma using a multiplex serological approach. Front Immunol. 8:1284. doi:10.3389/fimmu.2017.01284 PMID:29123514

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

Boffetta P, Gaborieau V, Nadon L, Parent MF, Weiderpass E, Siemiatycki J (2001). Exposure to titanium dioxide and risk of lung cancer in a population-based study from Montreal. Scand J Work Environ Health. 27(4):227–32. doi:10.5271/sjweh.609 PMID:11560336

Brøndum-Jacobsen P, Nordestgaard BG, Nielsen SF, Benn M (2013). Skin cancer as a marker of sun exposure associates with myocardial infarction, hip fracture and death from any cause. Int J Epidemiol. 42(5):1486–96. doi:10.1093/ije/dyt168 PMID:24038635

Cardis E, Deltour I, Vrijheid M, Combalot E, Moissonnier M, Tardy H, et al.; INTERPHONE Study Group (2010). Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. Int J Epidemiol. 39(3):675–94. doi:10.1093/ije/dyq079 PMID:20483835

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

Checkoway H, Pearce NE, Kriebel D (2004). Research methods in occupational epidemiology. 2nd ed. New York (NY), USA: Oxford University Press. doi:10.1093/acprof:oso/9780195092424.001.0001

Checkoway H, Pearce NE, Crawford-Brown DJ (1989). Research methods in occupational epidemiology. New York (NY), USA: Oxford University Press.

Danaei G, Tavakkoli M, Hernán MA (2012). Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. Am J Epidemiol. 175(4):250–62. doi:10.1093/aje/kwr301 PMID:22223710

Farr W, Humphreys NA (1885). Vital statistics: a memorial volume of selections from the reports and writings of William Farr. London, UK: The Sanitary Institute of Great Britain.

Flanders WD, Klein M (2007). Properties of 2 counterfactual effect definitions of a point exposure. Epidemiology. 18(4):453–60. doi:10.1097/01.ede.0000261472.07150.4f PMID:17473709

Fox MP, MacLehose RF, Lash TL (2021). Applying quantitative bias analysis to epidemiologic data. Statistics for biology and health. 2nd ed. Cham, Switzerland: Springer; doi:10.1007/978-3-030-82673-4_6

Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol. 186(9):1026–34. doi:10.1093/aje/kwx246 PMID:28641372

Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, et al. (2002). The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. Breast Cancer Res Treat. 74(3):235–54. doi:10.1023/A:1016387020854 PMID:12206514

Hanley JA, Foster BJ (2014). Avoiding blunders involving 'immortal time'. Int J Epidemiol. 43(3):949–61. doi:10.1093/ije/dyu105 PMID:24760815

Hernán MA (2015). Counterpoint: epidemiology to guide decision-making: moving away from practice-free research. Am J Epidemiol. 182(10):834–9. doi:10.1093/aje/kwv215 PMID:26507306

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, Sauer BC, Hernández-Díaz S, Platt R, Shrier I (2016). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 79:70–5. doi:10.1016/j.jclinepi.2016.04.014 PMID:27237061

Hill AB, Doll R (1956). Lung cancer and tobacco; the BMJ's questions answered. BMJ. 1(4976):1160–3. doi:10.1136/bmj.1.4976.1160 PMID:13316104

IARC (1997). Polychlorinated dibenzo-*para*-dioxins and polychlorinated dibenzofurans. IARC Monogr Eval Carcinog Risks Hum. 69:1–666. Available from: https://publications.iarc.who.int/87 PMID:9379504

IARC (2013). Malaria and some polyomaviruses (SV40, BK, JC, and Merkel cell viruses). IARC Monogr Eval Carcinog Risks Hum. 104:1–353. Available from: https://publications.iarc.who.int/128 PMID:26173303

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

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

INTERPHONE Study Group (2010). Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. Int J Epidemiol. 39(3):675–94. [Erratum in: Int J Epidemiol. 2012 Feb;41(1):328.] doi:10.1093/ije/dyq079 PMID:20483835

Jones ME, Schoemaker MJ, McFadden EC, Wright LB, Johns LE, Swerdlow AJ (2019). Night shift work and risk of breast cancer in women: the Generations Study cohort. Br J Cancer. 121(2):172–9. doi:10.1038/s41416-019-0485-7 PMID:31138896

Keil AP, Richardson DB, Troester MA (2015). Healthy worker survivor bias in the Colorado Plateau uranium miners cohort. Am J Epidemiol. 181(10):762–70. doi:10.1093/aje/kwu348 PMID:25837305

Kogevinas M, Becher H, Benn T, Bertazzi PA, Boffetta P, Bueno-de-Mesquita HB, et al. (1997). Cancer mortality in workers exposed to phenoxy herbicides, chlorophenols, and dioxins. An expanded and updated international cohort study. Am J Epidemiol. 145(12):1061–75. doi:10.1093/oxfordjournals.aje.a009069 PMID:9199536

Lange T, Keiding N (2014). Skin cancer as a marker of sun exposure: a case of serious immortality bias. Int J Epidemiol. 43(3):971. doi:10.1093/ije/dyu100 PMID:24894096

Mohebbi E, Hadji M, Rashidian H, Rezaianzadeh A, Marzban M, Haghdoost AA, et al. (2021). Opium use and the risk of head and neck squamous cell carcinoma. Int J Cancer. 148(5):1066–76. doi:10.1002/ijc.33289 PMID:32895947

Mukherjee Das A, Gogia A, Garg M, Elaiyaraja A, Arambam P, Mathur S, et al. (2022). Urinary concentration of endocrine-disrupting phthalates and breast cancer risk in Indian women: a case-control study with a focus on mutations in phthalate-responsive genes. Cancer Epidemiol. 79:102188. doi:10.1016/j.canep.2022.102188 PMID:35688051

Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G (2018). Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 47(1):226–35. doi:10.1093/ije/dyx206 PMID:29040562

Nasrollahzadeh D, Kamangar F, Aghcheli K, Sotoudeh M, Islami F, Abnet CC, et al. (2008). Opium, tobacco, and alcohol use in relation to oesophageal squamous cell carcinoma in a high-risk area of Iran. Br J Cancer. 98:1857–63. doi:10.1038/sj.bjc.6604369 PMID:18475303

O'Leary ES, Schoenfeld ER, Stevens RG, Kabat GC, Henderson K, Grimson R, et al.; Electromagnetic Fields and Breast Cancer on Long Island Study Group (2006). Shift work, light at night, and breast cancer on Long Island, New York. Am J Epidemiol. 164(4):358–66. doi:10.1093/aje/kwj211 PMID:16777931

Pearce N, Checkoway H, Dement J (1988). Exponential models for analyses of time-related factors, illustrated with asbestos textile worker mortality data. J Occup Med. 30(6):517–22. doi:10.1097/00043764-198806000-00015 PMID:3392615

Pearce N, Checkoway H, Shy C (1986). Time-related factors as potential confounders and effect modifiers in studies based on an occupational cohort. Scand J Work Environ Health. 12(2):97–107. doi:10.5271/sjweh.2161 PMID:3726500

Pearce N, Lawlor DA (2016). Causal inference – so much more than statistics. Int J Epidemiol. 45(6):1895–903. doi:10.1093/ije/dyw328 PMID:28204514

Pearce N, Richiardi L (2014). Commentary: three worlds collide: Berkson's bias, selection bias and collider bias. Int J Epidemiol. 43(2):521–4. doi:10.1093/ije/dyu025 PMID:24585732

Pearce NE, Smith AH, Fisher DO (1983). Ascertainment and correction of selection bias in an occupational cancer case-control study. Scand J Work Environ Health. 9:64.

Pinkerton LE, Hein MJ, Anderson JL, Little MP, Sigurdson AJ, Schubauer-Berigan MK (2016). Breast cancer incidence among female flight attendants: exposure-response analyses. Scand J Work Environ Health. 42(6):538–46. doi:10.5271/sjweh.3586 PMID:27551752

Pizzi C, De Stavola B, Merletti F, Bellocco R, dos Santos Silva I, Pearce N, et al. (2011). Sample selection and validity of exposure–disease association estimates in cohort studies. J Epidemiol Community Health. 65(5):407–11. doi:10.1136/jech.2009.107185 PMID:20881022

Richardson DB, Sugiyama H, Wing S, Sakata R, Grant E, Shimizu Y, et al. (2009). Positive associations between ionizing radiation and lymphoma mortality among men. Am J Epidemiol. 169(8):969–76. doi:10.1093/aje/kwp018 PMID:19270049

Richiardi L, Barone-Adesi F, Merletti F, Pearce N (2008). Using directed acyclic graphs to consider adjustment for socioeconomic status in occupational cancer studies. J Epidemiol Community Health. 62(7):e14. doi:10.1136/jech.2007.065581 PMID:18572430

Richiardi L, Pizzi C, Pearce N (2013). Commentary: representativeness is usually not necessary and often should be avoided. Int J Epidemiol. 42(4):1018–22. doi:10.1093/ije/dyt103 PMID:24062290

Saracci R, Kogevinas M, Bertazzi PA, Bueno de Mesquita BH, Coggon D, Green LM, et al. (1991). Cancer mortality in workers exposed to chlorophenoxy herbicides and chlorophenols. Lancet. 338(8774):1027–32. doi:10.1016/0140-6736(91)91898-5 PMID:1681353

Saracci R, Pearce N (2008). Commentary: observational studies may conceal a weakly elevated risk under the appearance of consistently reduced risks. Int J Epidemiol. 37(6):1313–5. doi:10.1093/ije/dyn200 PMID:18931393

Saracci R, Samet J (2010). Commentary: call me on my mobile phone...or better not?–a look at the INTERPHONE study results. Int J Epidemiol. 39(3):695–8. doi:10.1093/ije/dyq082 PMID:20483832

Schubauer-Berigan MK, Anderson JL, Hein MJ, Little MP, Sigurdson AJ, Pinkerton LE (2015). Breast cancer incidence in a cohort of US flight attendants. Am J Ind Med. 58(3):252–66. doi:10.1002/ajim.22419 PMID:25678455

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

Shakhssalim N, Hosseini SY, Basiri A, Eshrati B, Mazaheri M, Soleimanirahbar A (2010). Prominent bladder cancer risk factors in Iran. Asian Pac J Cancer Prev. 11(3):601–6. PMID:21039023

Sheikh M, Shakeri R, Poustchi H, Pourshams A, Etemadi A, Islami F, et al. (2020). Opium use and subsequent incidence of cancer: results from the Golestan Cohort Study. Lancet Glob Health. 8(5):e649–60. doi:10.1016/S2214-109X(20)30059-0

Smith AH, Pearce NE, Fisher DO, Giles HJ, Teague CA, Howard JK (1984). Soft tissue sarcoma and exposure to phenoxyherbicides and chlorophenols in New Zealand. J Natl Cancer Inst. 73(5):1111–7. PMID:6593487

Snoep JD, Morabia A, Hernández-Díaz S, Hernán MA, Vandenbroucke JP (2014). Commentary: a structural approach to Berkson's fallacy and a guide to a history of opinions about it. Int J Epidemiol. 43(2):515–21. doi:10.1093/ije/dyu026 PMID:24585735

Vandenbroucke J, Pearce N (2015a). Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? Am J Epidemiol. 182(10):826–33. doi:10.1093/aje/kwv225 PMID:26507305

Vandenbroucke J, Pearce N (2015b). Vandenbroucke and Pearce respond to "Incident and prevalent exposures and causal inference". Am J Epidemiol. 182(10):846–7. doi:10.1093/aje/kwv219 PMID:26507304

Vandenbroucke JP, Pearce N (2012). Case-control studies: basic concepts. Int J Epidemiol. 41(5):1480–9. doi:10.1093/ije/dys147 PMID:23045208

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

Wild CP (2005). Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 14(8):1847–50. doi:10.1158/1055-9965.EPI-05-0456 PMID:16103423

Xu M, Richardson L, Campbell S, Pintos J, Siemiatycki J (2018). Response rates in case-control studies of cancer by era of fieldwork and by characteristics of study design. Ann Epidemiol. 28(6):385–91. doi:10.1016/j.annepidem.2018.04.001 PMID:29703520

Chapter 6. Incorporating bias assessments into evidence synthesis

6.1 Introduction 160
6.2 Frameworks for incorporating bias assessment into evidence synthesis 161
6.3 Developing the bias-review process 162
6.4 Methods for studying multiple biases 167
6.5 Summary 172

CHAPTER 6.

Incorporating bias assessments into evidence synthesis

Amy Berrington de González, Nathan DeBono, Alexander P. Keil, Deborah A. Lawlor, Ruth M. Lunn, and David A. Savitz

6.1 Introduction

In the *IARC Monographs* programme, and in many other situations, experts are asked to examine, evaluate, and interpret a body of research that will then be used to make a judgement that could inform an authoritative statement, influence regulations, guide individual behaviours, or have other societal impact. In the *IARC Monographs* assessment process, the focus is on potentially preventable causes of cancer, but the same principles are applicable to other disease determinants and health outcomes. In public health, the determination of causation is rarely a simple yes–no decision. Rather, it requires the careful assembly of evidence and the use of inferential methods to reach a conclusion. Studies can provide information on the statistical relation between exposure and disease; by combining subject-matter expertise with an understanding of the study design and methods, considering complementary lines of research, and carefully examining the results, an assessment is made of the validity of the observed associations and their implications for inferences about causality. In almost all situations of interest, there will be more than one contributory study. The goal is to assess first the information value of each study, methodically and accurately, and then the totality of the available studies.

In this chapter, approaches are outlined for incorporating the wide array of bias assessment methods described in this book into the review process and evidence synthesis. This includes developing the process for the systematic review of key biases in individual studies and incorporating the bias assessment into the evidence synthesis. Two somewhat distinct approaches to the systematic review of biases are currently in use, which can be labelled as triangulation and algorithms. These two approaches are first described and contrasted, and then the rationale for a proposed third way is provided, drawing on the strengths of each. Three main steps in the bias-review process are outlined and illustrated with examples from the exposures used throughout this book. The chapter concludes with some discussion of methods for evaluating multiple sources of bias within a single study.

6.2 Frameworks for incorporating bias assessment into evidence synthesis

6.2.1 Triangulation

The triangulation of evidence from cancer epidemiology, animal bioassays, and mechanistic research is the overarching framework for the *IARC Monographs* review and classification system, as detailed in Chapter 1. Triangulation was introduced conceptually in Chapters 3 and 5 as a means of examining biases (specifically, confounding and selection bias) in individual studies. Triangulation can also serve as a framework for bias assessment across the epidemiological data. This approach emphasizes the benefits of examining the complete array of evidence to determine whether the varying strengths and limitations of the studies provide complementary information that helps in making an integrated assessment (Lawlor et al., 2016). The concept is particularly applicable when there is an array of studies with varying methodological strengths and limitations that could lead to bias in opposing directions.

Specifically, the aim in triangulation is to identify study designs (or approaches) that would be expected to have biases in opposing directions, to infer what a third, hypothetical, group of idealized studies would find. This inferred ideal can provide additional information about the probable bounds of a true causal effect. In practice, this can be implemented by contrasting studies through stratified meta-analysis or stratified forest plots. The approach requires consideration of the direction of the potential biases; this is an important strength. Example subgroups of studies that could be contrasted include the exposure setting, which might relate to the exposure level and degree of measurement error, for example studies of occupational versus environmental levels of exposure, cohort versus case–control study designs for assessment of recall bias, or cancer incidence versus mortality end-points for outcome misclassification. The study features should ideally involve complementary and exclusive biases that might affect one group of studies but not another. While no single study is likely to have all the desired positive features, a series of imperfect studies with complementary features could allow inference of what might be found in an ideal study.

By considering the full array of informative studies, there is an emphasis on corroboration, which links back to Hill's viewpoint on consistency of findings across a variety of locations and populations (Hill, 1965) and is consistent with the *IARC Monographs* Preamble (IARC, 2019), as noted in Chapter 1. Triangulation emphasizes the exploration of sources of heterogeneity. The reasoning is logical, intuitive, and flexible in being adaptable to a range of topics and diverse methodological concerns. These are all additional strengths. A weakness, currently, is that triangulation is a broad rather than a specific approach, lacking standardization; this could be invoked as a rationale that leads to a range of conclusions. In drawing on subject-matter expertise to interpret a given set of studies, there is latitude in what is emphasized and what is downplayed.

6.2.2 Algorithms

Several algorithms for bias assessment in epidemiology have been proposed, including Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) (Sterne et al., 2016) or of Exposure (ROBINS-E) (Higgins et al., 2022, 2024) and risk-of-bias scales, such as the Newcastle–Ottawa instrument (Deeks et al., 2003). A strength is that they offer a comprehensive set of rules and procedures to follow, with the intent of providing an evaluation that is replicable and objective and can be conducted by non-experts. A concern is the unwarranted degree of confidence that the algorithm gives the so-called correct answer (Igelström et al., 2021). There is no gold standard to know when an answer is right or wrong, and it is preferable to acknowledge the complexity of inferences about causality and accept the burden of explaining the reasoning that leads to the judgement, instead of simply invoking an algorithmic methodology. The aspiration of eliminating the subjectivity of reviewers and ensuring replicability is laudable, but it is unrealistic to expect that a generic algorithm for judging study quality will apply with equal validity to all exposures and all outcomes.

The comprehensive nature of the current algorithms, often involving a lengthy series of questions covering every potential source of bias, can also be a weakness. If there is no initial evaluation by subject-matter experts of the domains that are key or influential biases for the exposure and outcome of interest, then the application of the algorithm to every study tends to pare down the evidence that is used, with studies accepted or rejected due to possibly

minor or misplaced concerns rather than acknowledging that each has strengths and limitations. Many algorithms also do not emphasize an evaluation of the direction of the bias. In hazard identification, the direction is especially important. There is the potential for substantial loss of information about a potential hazard if all positive studies with bias towards the null were excluded, for example. The ability to assess consistency and the role of chance is also reduced if only a small subset of studies is retained.

These algorithms are often used in conjunction with the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) framework (Guyatt et al., 2008), developed for assessing clinical or other forms of experimental research, which automatically downgrades the value of observational studies in the evidence synthesis. Randomization is rarely ethical or feasible with etiological studies of cancer (other than prevention trials) and often requires the forfeit of other important study attributes, including exposure range, prolonged exposures, and study size. The strengths and weaknesses of different study designs will depend on the specific exposure and outcome under consideration.

Finally, although an algorithm may be presented as well-defined and systematic, there is still abundant opportunity to have the opinions of those implementing it influence the outcome. To the extent that there is a need for subject-matter expertise and an inherent intrusion of individual judgements, it is preferable to present the fact transparently rather than to mask it behind an algorithm.

6.2.3 Concluding thoughts about frameworks

There are strengths and weaknesses of triangulation and algorithms, as currently proposed, as bias assessment frameworks for epidemiological studies. A third way lies between the rigid approach of algorithms and the general approach of triangulation. This third way involves laying out a bias assessment process for the specific exposure and outcome under review that uses the full array of informative studies and the wide array of tools described in this book to assess the direction and magnitude of potential biases. This proposed approach is consistent with the review methods described in the *IARC Monographs* Preamble (IARC, 2019; see Chapter 1), which calls for Working Groups to integrate studies in evidence synthesis on the basis of their quality and informativeness but recommends against the use of checklists to assess biases and sources of error. It is also recommended that the bias assessment process be led by subject-matter experts, including epidemiologists, statisticians, and exposure assessors, again consistent with the *IARC Monographs* assessments. The following sections outline the key steps in this proposed approach and illustrate it with examples.

6.3 Developing the bias-review process

A bias-review process, developed by subject-matter experts, can guide the systematic review of biases in each individual study and at the evidence synthesis stage. There are typically several steps in the process, as outlined in Fig. 6.1: (i) a definition of the key biases for the exposure–outcome under consideration; (ii) a review, and a summary, of the informative studies for these key biases; and (iii) an assessment of the influence of the key biases on the study findings. The process is specific to each pair of exposures and outcomes under consideration and can be iterative. For the *IARC Monographs* evaluations, the Preamble and instructions for authors provide a starting point, and substance-specific issues can be added to the meeting-specific instructions for authors (IARC, 2024). These steps are described in more detail next.

6.3.1 Determining the key types of bias

A key step for the expert review group is to consider which of the many potential biases are of greatest concern. This will depend on the specific exposure–cancer outcome pair under review, and on the types of study that are available. A directed acyclic graph (DAG), as described in Chapter 2, can help the expert reviewers to reach agreement on the possible bias domains. Once these bias domains are agreed on, some specific signalling questions can be developed to guide the reviewers in their considerations. These questions should help identify the direction and likely magnitude of the bias, not simply its presence or absence. Chapters 3, 4, and 5 can help the reviewers make these determinations. Deciding which biases are not relevant, or not likely to be material, helps to focus the reviewers' attention on the critical subset. This process is illustrated in Examples 6.1 and 6.2.

Fig. 6.1. Steps in the bias-review process.

Example 6.1. Selection of key biases for night shift work

Because night shift work is a complex exposure scenario, the *IARC Monographs* Working Group stated in its assessment of the evidence in humans that "exposure assessment quality of night shift work was a key parameter for the evaluation of the studies" (IARC, 2020), and the reviewers conducted an extensive evaluation of this aspect of each study. In contrast, the Working Group noted that although differences in lifestyle factors exist between day and night shift workers, these differences are usually small; this suggests that the reviewers considered confounding to be of lesser concern. Because there were many informative case–control studies, which tended to have more detailed exposure assessment, selection bias was examined, along with recall bias. (text continues on page 162)

Example 6.2. Selection of key biases for opium consumption

There were a wide range of concerns about potential biases in the epidemiological studies of opium consumption, and the *IARC Monographs* Working Group documented its considerations in an annex to *IARC Monographs* Volume 126 (IARC, 2021), which serves as an example of a bias assessment framework. The Working Group noted that key potential biases for the examined studies of opium consumption included reverse causation (consumption of opium because of a cancer diagnosis) and protopathic bias (consumption of opium to alleviate prediagnostic symptoms). In addition, there were concerns about selection bias because there were several hospital-based case–control studies. Non-differential exposure misclassification and inclusion of infrequent opium users in the baseline category used for exposure–response analyses were thought to lead to downward bias. Finally, there were other strong risk factors for the cancers under study, particularly tobacco use, which had been shown in the exposure assessment review to be strongly related to opium use; thus, confounding was also a potential bias. (text continues on page 164)

(a) Guidance for identifying key confounders

Once the key bias domains have been identified for the specific exposure–cancer scenario under investigation, the review team should provide additional details to guide the bias assessment. For confounding, the Working Group members should use their expertise and literature reviews to identify all the key confounders, i.e. those variables most likely to bias the effect estimate and distort its interpretation if they are not controlled for in the study. The use of DAGs can guide and help document these decisions (see Chapter 2). An approach to this identification is given in Side Box 6.1 and Example 6.3. The methods in Chapter 3 can help in assessing the likely direction and magnitude of confounding. There should also be consideration of whether certain variables could be effect modifiers or mediators, rather than confounders, because adjustment for these could introduce, rather than remove, bias (see Chapter 2 for more details).

(b) Guidance for assessing misclassification and mismeasurement of exposure

In general, the bias framework for exposure misclassification should cover how well the exposure proxy approximates the exposure of interest, the extent of measurement error, and whether the measurement error is differential or non-differential. Side Box 6.2 lists scoping questions to inform the bias evaluation. The methods described in Chapter 4 can help to determine the likely direction and possible magnitude of bias from misclassification and mismeasurement of exposure, as illustrated in Example 6.4.

(c) Guidance for assessing other key biases

The detailed guidance described above for confounding and measurement error provides examples of thorough assessment of the key concerns for these topics. Other topics may call for analogous assessments of other types of bias, for example selection bias, healthy worker effects, and outcome misclassification. For each key bias, a set of questions should be identified and guidance provided. For example, for selection bias, reviewers should consider sources of bias such as study inclusion or exclusion criteria, sources of control participants, and missing data or loss to follow-up (Example 6.5). It may help to use DAGs to illustrate sources of selection bias, including colliders. For a detailed evaluation of how to identify selection bias in case–control studies, see Section 5.3.

6.3.2 Summarizing the bias assessment and synthesizing across studies

A table summarizing the results from the review of the key biases in each study is recommended. For instance, in Example 6.6, for an analysis of studies on opium consumption and

Side Box 6.1. Approach for identifying key confounders

(i) Identify the known causes of the cancer (e.g. those with *sufficient* or *limited* evidence of causality) by consulting experts with relevant subject knowledge and using authoritative sources, such as the *IARC Monographs* and the *IARC Handbooks of Cancer Prevention*, the United States National Toxicology Program Report on Carcinogens, and the World Cancer Research Fund. Specify the hypothesized direction of the confounder–cancer association (e.g. relative risk [RR] > 1 or RR < 1).

(ii) Identify which cancer causes are plausibly related to the exposure of interest, by using authoritative sources and consulting experts with relevant subject knowledge. This information is often reported in the section on exposure characterization of the relevant *IARC Monograph*. Specify the hypothesized direction of the confounder–exposure association (e.g. RR > 1 or RR < 1).

(iii) Research (e.g. conduct literature searches, seek expert opinion on mechanistic data) whether the identified potential confounders could be mediatory (in the causal pathway between the exposure and cancer) rather than confounders. It may be helpful to construct a DAG to identify mediators and colliders, which should not be controlled for in studies.

(iv) Identify the minimal set of key variables necessary to control for confounding, and assess the expected direction of the bias (the methods outlined in Chapters 2 and 3 can be helpful). (text continues above)

Example 6.3. Specifying key confounders

Returning to the example of night shift work in relation to breast cancer, this example illustrates how the approach outlined in Side Box 6.1 can be used to specify key confounders.

Table 6.1 lists the causes of female breast cancer identified from IARC, the World Cancer Research Fund, the United States National Toxicology Program Report on Carcinogens, and literature reviews, and the subset of these that could be considered as potential key confounders for night shift work. Age at first full-term pregnancy could be considered the key confounder for reproductive breast cancer factors because other factors, such as parity, are often related to it, and some of their confounding effects are likely to be controlled for by controlling for age at first full-term pregnancy. Other pharmacological and lifestyle factors, such as the use of oral contraceptives and tobacco smoking, might not be key confounders because of relatively weak associations with breast cancer. In contrast, although a family history of breast cancer is strongly associated with breast cancer risk, it would be unlikely to be associated with night shift work and would therefore not be a key confounder. Occupational exposure to ionizing (cosmic) radiation could be a key confounder in flight crew studies because of its high correlation with night work hours. (text continues on page 164)

Table 6.1. Potential key confounders for night shift work and female breast cancer

Potential confounding factors	Causes of female breast cancer[a]	Key confounders (and expected directions)
Reproductive and family history factors	Early age at menarche, late age at first full-term pregnancy, nulliparity, menopausal status or age at menopause, no breastfeeding, family history of breast cancer	Young age at first full-term pregnancy or parity. These are protective for breast cancer and are probably negatively associated with night shift work; therefore, confounding away from the null.
Lifestyle factors	Lack of physical activity (primarily postmenopausal breast cancer), obesity (increases risk in postmenopausal women; decreases risk in premenopausal women), consumption of alcoholic beverages, tobacco smoking	Obesity is a risk factor for breast cancer and is probably positively associated with night shift work; therefore, bias probably away from the null. Note that obesity could be a mediator or a confounder (or both).
Pharmacological factors	Diethylstilbestrol, estrogen–progestogen contraceptives, hormone menopausal therapy (estrogen–progestogen or estrogen only), digoxin	
Demographics	Age, socioeconomic status, education level	Age, socioeconomic status, and education level are negatively associated with breast cancer, but the direction of the association with night shift work depends on the profession.
Occupational agents	X-radiation, gamma radiation, ethylene oxide, polychlorinated biphenyls	Cosmic radiation for aircrew workers. The direction of the bias depends on the comparison group (e.g. day workers or non-workers) and the study population.

[a] NTP (2018).

Side Box 6.2. Approach for assessing exposure misclassification and measurement error

- How was exposure assessed in the studies under review (e.g. questions, records, environmental measurements, biomarkers)?
- The Working Group should research the following questions for each type of exposure assessment.
  - What was the temporal sequence of exposure and outcome measurement? Could disease status have affected the exposure measurement?
  - Are there methods that are prone to major error or are biologically inappropriate for the exposure of interest? For example, a biomarker with a short half-life might not be informative for evaluating cancer risk.
  - What are the ideal methods for evaluating exposure?
  - What are the potential sources of measurement error?
  - What is the type of measurement error (e.g. classical, Berkson, differential, non-differential)?
  - Are there validation studies available? What values of sensitivity and specificity do the validation studies report? (text continues on page 164)

Example 6.4. Assessing exposure misclassification

For studies on night shift work and breast cancer, the most common methods to assess and classify exposure involved using questionnaires, payroll records, or a population-based job-exposure matrix (e.g. based on survey data reporting the percentage of night shift workers for different job categories). Table 6.2 lists questions and considerations for assessing the potential biases from exposure misclassification in the studies on night shift work. (text continues on page 164)

Table 6.2. Assessment of exposure misclassification for studies on night shift work

Questions	Guidance, comments
What is the source of the exposure assessment? Is there concern that the exposure assessment did not distinguish between exposed and non-exposed people or among exposure categories during a relevant time window of exposure? What are the likely direction and magnitude of bias?	Questionnaires, interviews: Ideally, the questionnaire should cover actual hours worked and lifetime work history. The group defined as unexposed might be exposed (i.e. previous night shift work) in studies ascertaining current exposure. Ideally, in cohort studies, information should be collected after baseline. Job-exposure matrix (JEM): Population-based JEMs are less informative than industry-specific JEMs. Information (e.g. census data) of the proportion of night shift workers in the same geographical region as the study population could provide some indication of the quality of the data. Payroll records: These are objective but are usually not complete because industry-specific records do not capture lifetime exposure from jobs in other workplaces. In general, bias from lower-quality exposure assessment is likely to be non-differential and towards the null (see Chapter 4 for exceptions).
Is there concern about differential recall?	Differential recall bias (most likely away from the null) is a potential concern in case–control studies. It may be less likely in studies published before the 2007 IARC evaluation of shift work (IARC, 2010). External research may help inform the assessment of recall bias (see Chapter 5 for examples).

Example 6.5. Identifying selection bias

Table 6.3 illustrates how this approach to assessing the potential for selection bias can be applied to a bias-review framework for case–control studies on opium consumption and various cancers. (text continues on page 164)

Table 6.3. Identifying selection bias for case–control studies on opium consumption

Question	Guidance, comments
Is there concern that selection into (or out of) the study was related to both exposure and outcome, and what is the likely direction of the bias?	Hospital control participants: Potential bias downwards if opium use is related to hospitalization and hospital control participants are more likely than the general population to have used opium.
	Neighbourhood control participants: Potential bias upwards if control participants who use opium are less likely to participate (e.g. leading to a lower exposure prevalence).

bladder cancer, Table 6.4 has one row per study and a column for each key bias, and gives the likely direction of potential bias. It is then easy to see the biases that have been identified for each study, and the groups of studies that have been identified with a certain bias. The table can be used to inform a triangulation process by identifying subsets of studies with differing key sources of biases, particularly where some studies would be expected to produce bias in opposing directions. The table also shows whether biases cluster within subsets of studies; this might make it difficult to separate the impact of specific biases. When multiple key biases affect a study, assessment of the total (resultant) bias is non-trivial. Section 6.4 describes some approaches and the related challenges involved in assessing multiple biases within a single study.

The extent to which it is then feasible to integrate study results and bias assessment can be influenced by how many informative human studies have been identified for review. For triangulation, the aim is to compare results from at least two, but ideally more, studies that have different key sources of bias. The study results can be contrasted via stratified forest plots or, more formally, by means of stratified meta-analysis to explore the impact of the bias. When there are many informative studies, the opportunities for bias assessment through triangulation are increased, particularly if there is a variety of settings and study designs. The different steps in the process are illustrated with examples including studies on opium consumption and bladder cancer (Example 6.6), a situation with only a few informative studies (mobile phone use and glioma; Example 6.7), and a quantitative triangulation of meta-analysis results where there are a large number of studies (red meat consumption and colorectal cancer; Example 6.8).

6.4 Methods for studying multiple biases

As seen in the examples throughout this book, studies could be subject to multiple key biases. At the evidence synthesis stage, the reviewers will then need to consider what the combined effect of those biases might be, and whether the combination could alter the interpretation. To answer this question requires consideration of the magnitude of each bias, along with the direction of the bias and some understanding of whether the biases act independently. This section discusses the issues that need to be considered when assessing the likely impact of multiple biases, how to approach multiple-bias sensitivity analyses, and when an individual-level data reanalysis could be important. Annex 3 includes a worked example of a formal multiple-bias analysis for a study on opium consumption and bladder cancer, which illustrates the complexity and the need to specify multiple parameters. Because of the

Example 6.6. Bias assessment summary table

In the review of the human evidence for *IARC Monographs* Volume 126 on opium consumption and bladder cancer, one cohort study and several case–control studies were considered informative (IARC, 2021). As noted in Section 6.3.1, there were a considerable number of potential key biases, which were discussed in an annex, titled "Methodological considerations for epidemiological studies on opium consumption and cancer" (IARC, 2021). A meta-analysis published subsequently used the bias assessment to explore between-study heterogeneity; that assessment is used here to illustrate how the biases can be summarized and synthesized (Miranda Filho et al., 2023).

Table 6.4 shows that most studies were not considered to be at risk of material (major) confounding bias or reverse causation, but that many of the case–control studies were considered to be at risk of selection bias and information bias. The direction of selection bias was identified as likely downwards in several hospital-based case–control studies, but of uncertain direction in others. The potential for recall bias and exposure misclassification was considered quite low, but these biases could operate in different directions, hence the arrow showing that this could result in bias towards or away from the null. In all the studies, a positive association was found between opium consumption and bladder cancer, but the magnitude of risk for ever or never having used opium varied widely, with an odds ratio of 2.47 to 8.23 and a summary estimate of 4.07 (95% confidence interval [CI], 3.23–5.12). Miranda Filho et al. (2023) conducted sensitivity analyses by excluding studies with various biases (e.g. selection bias, information bias). The summary relative risk was slightly lower in the studies considered to have low risk of selection bias (odds ratio [OR], 3.40; 95% CI, 2.70–4.30) or information bias (OR, 3.69; 95% CI, 3.01–4.41) but was still strongly supportive of a positive association.

Table 6.4. Bias assessment summary for studies on opium consumption (ever vs never use) and bladder cancer based on major concerns, as defined and identified by Miranda Filho et al. (2023)[a]

Study (first author)	OR or RR (CI)[b]	Design	Confounding	Reverse causation	Selection bias	Information bias	Protopathic bias
Sheikh	2.86 (1.47–5.56)	co					
Aliasgari	2.60 (0.80–8.47)	c–c(h)			←	↔	
Aliramaji	4.10 (1.59–10.55)	c–c(h)			←	↔	
Sadeghi	2.70 (0.18–40.81)	c–c(h)			←	↔	
Nourbakhsh	3.87 (1.98–7.57)	c–c			↔	↔	
Tootoonchi	2.45 (0.98–6.14)	c–c			↔	↔	
Abdolahinia	8.23 (3.82–17.71)	c–c			↔	↔	
Akbari	3.90 (1.28–11.85)	c–c					
Hadji	3.40 (2.69–4.29)	c–c					
Rashidian	4.40 (2.94–6.59)	c–c					
Ghadimi	4.96 (1.07–22.96)	c–c(h)			←	↔	
Hosseini	4.16 (2.67–6.47)	c–c(h)			←		
Ketabchi	7.99 (5.20–12.27)	c–c			↔	↔	
Lofti	3.01 (1.73–5.23)	c–c				↔	
Shakhssalim	2.57 (1.55–4.26)	c–c				↔	

c–c, case–control; c–c(h), hospital-based case–control; CI, confidence interval; co, cohort; OR, odds ratio; RR, relative risk.

[a] Arrows indicate the direction of the biases: ←, downwards; ↔, uncertain direction. Blank indicates that the reviewers concluded that there was no substantial bias.

[b] Controlling for tobacco smoking, where available.

Example 6.6. Bias assessment summary table (continued)

The Working Group concluded that chance, bias, and confounding could be ruled out because of the strong associations and the consistency across studies and across the study designs (e.g. the cohort and the case–control studies with different sources of control participants). The Working Group did not use the term *triangulation* but commented, "It is notable that the results of all studies, regardless of design, point in the same direction" (IARC, 2021). (text continues on page 167)

Example 6.7. Bias assessment summary with few informative studies

In a review of studies on radiofrequency electromagnetic field radiation exposure (mainly through mobile phone use) and brain tumours (IARC, 2013), the reviewers considered most of the early small case–control studies to be relatively uninformative. Therefore, evaluation of the human evidence was based largely on two large case–control studies: the Interphone multicentre case–control study (Cardis et al., 2011) and a large case–control study in Sweden (Hardell et al., 2011).

In the case–control study in Sweden, with 1148 cases of glioma and 2438 control participants, Hardell et al. (2011) reported a monotonically increasing risk of glioma with increasing cumulative duration of mobile phone use, with an odds ratio of 3.2 (95% CI, 2.0–5.1) for > 2000 hours use compared with no use. In the Interphone study, with 2708 cases of glioma and 2792 control participants, cumulative call time was divided into deciles, with a referent comprising those who had never regularly used mobile phones. In contrast to the findings from the case–control study in Sweden, in the Interphone study, the odds ratios were mostly < 1 (ranging from 0.7 to 1.05), except for the highest category, of ≥ 1640 hours of cumulative call time (OR, 1.40; 95% CI, 1.03–1.89). Because these were case–control studies based on self-reported mobile phone use, the review group identified differential measurement error (recall bias) and selection bias as the key potential sources of bias. Because there are few established risk factors for glioma, confounding was considered less of an identifiable problem.

Selection bias was of greater potential concern in the Interphone study, because the participation rates were relatively low, especially for control participants (64% for cases and 53% for controls). In the case–control study in Sweden, participation rates were higher and non-differential (85% for case participants and 84% for control participants). In the Interphone study, a short non-response questionnaire revealed that the participation rate was higher in regular mobile phone users, particularly for case participants. When the analysis was restricted to regular users (i.e. by changing the reference category), the odds ratios for cumulative call time changed qualitatively to become mostly > 1 (increasing by 20–50%). Although there was still no clear evidence of a dose–response relation across the 10 categories of duration, the odds ratio for ≥ 1640 hours of cumulative call time increased from 1.40 (95% CI, 1.03–1.89) to 1.82 (95% CI, 1.15–2.89).

There were also extensive efforts to evaluate the quality of the exposure data in the Interphone study; these included a substudy with software-modified phones and phone records, which found substantial reporting error, with some indication of greater overreporting by case participants (Vrijheid et al., 2006, 2009). Exclusion of all participants who reported usage for > 5 hours per day decreased the odds ratio in the highest decile from 1.40 to 1.27 (95% CI, 0.92–1.74), but truncation at 5 hours per day did not influence the odds ratio. As explained in Chapter 4, bias from non-differential misclassification in categorical variables is not necessarily towards the null. Table 6.5 summarizes the bias assessment for the key domains, and the likely direction of the bias for the two informative studies.

Example 6.7. Bias assessment summary with few informative studies (continued)

Table 6.5. Bias assessment summary for case–control studies on mobile phone use and glioma[a]

Study	Risk estimate (95% CI)	Information bias	Selection bias
Case–control study in Sweden (Hardell et al., 2011)	3.2 (2.0–5.1)[b]	↔	
Interphone study (Cardis et al., 2011)	1.40 (1.03–1.89)[c]	↔	←

CI, confidence interval.

[a] Arrows indicate the direction of the biases: ←, downwards; ↔, uncertain direction. Blank indicates that the reviewers concluded that there was no substantial bias.

[b] Highest exposure category of > 2000 hours of cumulative call time.

[c] Highest exposure category of ≥ 1640 hours of cumulative call time.

Because there were only two informative studies and they had a similar design (population-based case–control studies with self-reported mobile phone use), triangulation was not possible. The higher risk of selection bias in the Interphone study, with some evidence that this was biased downwards, could partly explain the difference in the magnitude of the risk estimates for the highest exposure category. However, these studies share the limitation of potential for recall bias and exposure misclassification, which could have opposing directions. Therefore, the assessment of the human evidence by the committee was that although there was a positive association between mobile phone use and the incidence of glioma, chance, bias, or confounding could not be ruled out with reasonable confidence. Multiple-bias analysis could have been used to further explore the combined effect of these biases. As noted in Section 6.4, this is a complex task and involves several assumptions and specification of multiple-bias parameters but can provide bounds on the plausible range of results. (text continues on page 167)

effort involved, it is worth considering whether multiple-bias assessment is necessary. For example, if the study result is positive and all key biases are expected to be towards the null, it is unnecessary to carry out a formal multiple-bias analysis for hazard identification.

There are two different approaches for sensitivity analysis: bias-level sensitivity analysis and target-adjusted sensitivity analysis. In bias-level sensitivity analysis, plausible bias values and structures are used to identify a range of results that the study could have obtained. When dealing with multiple biases, the order of corrections must be considered. For example, should one adjust for confounding or exposure misclassification first? Fox et al. (2021) recommend what they term sequential bias analysis, in which biases are adjusted for sequentially in the reverse order of which they likely occurred. A common sequence in which biases arise would be confounding, followed by selection bias, and finally exposure misclassification, but this is not always the case. The order of analysis matters because sensitivity and specificity parameters, for example, may differ, depending on whether misclassification of the exposure or outcome occurs before or after study selection (Example 6.9). Ross et al. (2022) show how adjusting for biases in the wrong order using individual-level data can lead to misadjustment and residual bias.

Example 6.9 highlights the challenges of conducting a multiple-bias analysis, of which there are very few examples in the literature. If evidence hinges on a single study in which multiple biases are suspected, such an analysis may be informative, but it should be interpreted cautiously, because of the inherent dependence on the accuracy of bias parameters. Probabilistic bias analysis accounts for uncertainty in the bias parameters and is discussed in Chapters 4 and 5. This uncertainty is quantified by proposing a distribution, rather than a single value, for each bias parameter. At the extreme end of probabilistic bias analysis is bounding, which involves finding the largest amount of bias that could result from the plausible distribution of bias parameters. In principle, a bounding approach can help to answer questions about

Example 6.8. Bias assessment summary using triangulation

This example is an illustration of triangulation using meta-analyses of studies of the association between red meat consumption and colorectal cancer. Results are stratified according to study designs that are likely to have biases in opposing directions: cohort and case–control. It is assumed that non-differential exposure misclassification is a source of bias towards the null in cohort studies with a single dietary questionnaire of limited detail. Also, it is assumed that recall bias is away from the null in the case–control studies, for example through case participants overreporting their exposure because of their diagnosis. These biases are unrelated, in that each bias affects one group of studies (i.e. cohort studies, case–control studies) but not the other.

From the results reported by Norat et al. (2002), the meta-effect estimate (Table 6.6) for the highest versus lowest quantile of consumption from the cohort studies (1.27) is slightly lower than that from the case–control studies (1.36). It is then possible to make inferences about a third, hypothetical, meta-effect estimate from an idealized study with no biases. Triangulation of the stratum-specific effect estimates suggests that the true causal effect may be between these two values. In this way, a bounded range of the magnitude of the causal effect is obtained, using information from two groups of studies. This approach is likely to be more informative than making inferences from one group of studies, because of a perceived methodological strength, while ignoring another. The IARC Working Group also identified several key confounders for the association between red meat consumption and colorectal cancer, including total energy (caloric) intake, physical activity, smoking, and body mass index. Stratified meta-analyses based on the degree of control for confounding within the subsets of case–control and cohort studies could provide further insight into the impact of confounding, and potential mediation for body mass index (as discussed in Example 2.1a). Additional insights into the potential impact of measurement error in red meat consumption are also shown in Example 4.22, which illustrates regression calibration. Calibration corrections of this type can be important in meta-analyses because they can reduce an important source of heterogeneity in effect estimates. (text continues on page 167)

Table 6.6. Example triangulation exercise, comparing meta-analysis results from studies of the association between red meat consumption and colorectal cancer

Strata	Source of bias	Direction of bias	Number of studies[a]	Meta-effect estimate (95% CI)[a]	Triangulated meta-effect estimate
Cohort studies	Non-differential exposure misclassification	Towards the null	9	1.27 (1.11–1.45)	1.27–1.36
Case–control studies	Recall bias	Away from the null	14	1.36 (1.17–1.59)	

CI, confidence interval.

[a] Results from Norat et al. (2002) for the highest quantile of red meat consumption.

Example 6.9. Multiple-bias analysis

A multiple-bias analysis within a systematic review may be most usefully undertaken when one has to consider a study of moderate to large size and evidence is uncertain. The study by Aliramaji et al. (2015) was one of the largest conducted to examine the relation between opium consumption and bladder cancer, and it was suspected to suffer from multiple key biases, whose directions might have offset each other (Table 6.4). In the original study publication, a crude odds ratio of 2.7 was reported for the opium consumption–bladder cancer association. Multiple biases were likely in this study. First, frequency matching on sex without adjustment would have introduced selection bias. Second, there was concern about exposure misclassification, because of the illicit nature of opium use. Third, there was a potential for uncontrolled confounding by smoking and sex (which are not noted in Table 6.4 because adjusted estimates were used for that determination, rather than the crude estimate reported by Aliramaji et al., 2015). Bias parameters to adjust for each of these biases were drawn from various sources, including survey data and a validation study of recent opium use conducted for a hospital-based cohort. A limitation of the bias analyses is the lack of validation studies of long-term opium use, which would have yielded misclassification parameters for long-term use. Instead, bias parameters were drawn from studies of recent use, which were available because there are reliable biomarkers of recent opium exposure.

The use of matching on sex probably led to downward selection bias because it resulted in an oversampling of men, who were less likely to be unexposed control participants. However, there was also uncontrolled confounding by sex, which was considered to be upwards. Adjustment for all three biases in the reverse order of which they were expected to occur (here the adjustment order was selection bias, exposure misclassification, and confounding) yielded an adjusted summary odds ratio of 8.6, suggesting that the bias in the study by Aliramaji et al. (2015), given the best available estimates of bias parameters, was downwards. This large change occurred because most adjustments were in the same direction. A full probabilistic bias analysis was not straightforward, given the studies from which the bias parameters were drawn. To address this partially, a sensitivity analysis was performed, in which the sensitivity parameters for exposure misclassification were varied within a relatively narrow range of plausible values. Even with this narrow range, the adjusted odds ratio ranged from 1.3 (no misclassification) to 11.2 (differential misclassification); this emphasizes the potential influence of this source of bias. Full details of this example are provided in Annex 3, along with R code (online only; available from: https://publications.iarc.who.int/634#supmat). (text continues on page 170)

whether an observed association might be due to bias alone. However, in practice, this bounding approach has not been widely used for multiple-bias analysis and is limited when considering biases that might offset each other. Finally, target-adjusted sensitivity analysis, such as the E-value (described in Section 3.3.4(c), which outlines a modified approach), involves identifying the extent of bias necessary for a given study result to be compatible with the null (or another) hypothesis. This approach would also likely be very difficult in a multiple-bias analysis using only published data and would involve unrealistic assumptions. Smith et al. (2021) give an example based on individual subject data.

6.5 Summary

This chapter provides pragmatic guidance on the development and application of bias assessment as part of the evidence synthesis process. A third way is offered, which lies between the rigid approach of algorithms and the general approach of triangulation. A critical philosophical distinction from the algorithmic approach is that this bias assessment should be developed and applied by multidisciplinary experts. This expertise facilitates the identification of key sources of bias for the specific exposure–cancer relation under review. Focusing on key sources of bias facilitates the review process and avoids the elimination of informative studies due to minor biases or biases that do not change the causal interpretation.

It is recommended to retain all informative studies and to document

the potential key biases, including their direction and, if possible, their magnitude. The array of studies can then be used to evaluate biases indirectly, and to triangulate epidemiological evidence by comparing results from subsets of studies with different key biases. The wide array of tools described in this book provides methods to evaluate the direction and magnitude of bias, drawing on external data where necessary. It is hoped that, in the future, these bias analyses will be incorporated into the results section in more original study publications, as outlined in Chapter 7, reducing the need for speculation in the ubiquitous paragraph on strengths and limitations in the discussion section of publications. This will strengthen the field of epidemiology and facilitate the bias assessment work of review teams.

References

Aliramaji A, Kaseean A, Yousefnia Pasha YR, Shafi H, Kamali S, Safari M, et al. (2015). Age distribution types of bladder cancers and their relationship with opium consumption and smoking. Caspian J Intern Med. 6(2):82–6. PMID:26221505

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

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al.; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group (2003). Evaluating non-randomised intervention studies. Health Technol Assess. 7(27):iii–x, 1–173. doi:10.3310/hta7270 PMID:14499048

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

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al.; GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 336(7650):924–6. doi:10.1136/bmj.39489.470347.AD PMID:18436948

Hardell L, Carlberg M, Hansson Mild K, Eriksson M (2011). Case-control study on the use of mobile and cordless phones and the risk for malignant melanoma in the head and neck region. Pathophysiology. 18(4):325–33. doi:10.1016/j.pathophys.2011.06.001 PMID:21764571

Higgins J, Morgan R, Rooney A, Taylor K, Thayer K, Silva R, et al. (ROBINS-E Development Group) (2022). Risk of bias in non-randomized studies - of exposure. Launch version, 1 June 2022. Available from: https://www.riskofbias.info/welcome/robins-e-tool.

Higgins JPT, Morgan RL, Rooney AA, Taylor KW, Thayer KA, Silva RA, et al. (2024). A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int. 186:108602. doi:10.1016/j.envint.2024.108602 PMID:38555664

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300. doi:10.1177/003591576505800503 PMID:14283879

IARC (2010). Painting, firefighting, and shiftwork. IARC Monogr Eval Carcinog Risks Hum. 98:1–804. Available from: https://publications.iarc.who.int/116 PMID:21381544

IARC (2013). Non-ionizing radiation, Part 2: Radiofrequency electromagnetic fields. IARC Monogr Eval Carcinog Risks Hum. 102:1–462. Available from: https://publications.iarc.who.int/126 PMID:24772662

IARC (2019). Preamble to the *IARC Monographs* (amended January 2019). Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

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

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

IARC (2024). *IARC Monographs* Preamble – Instructions for authors. Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/preamble-instructions-for-authors/.

Igelström E, Campbell M, Craig P, Katikireddi SV (2021). Cochrane's risk of bias tool for non-randomized studies (ROBINS-I) is frequently misapplied: a methodological systematic review. J Clin Epidemiol. 140:22–32. doi:10.1016/j.jclinepi.2021.08.022 PMID:34437948

Lawlor DA, Tilling K, Davey Smith G (2016). Triangulation in aetiological epidemiology. Int J Epidemiol. 45(6):1866–86.doi:10.1093/ije/dyw314 PMID:28108528

Miranda Filho A, Turner MC, Warnakulasuriya S, Richardson DB, Hosseini B, Kamangar F, et al. (2023). The carcinogenicity of opium consumption: a systematic review and meta-analysis. Eur J Epidemiol. doi:10.1007/s10654-023-00969-7 PMID:36773182

Norat T, Lukanova A, Ferrari P, Riboli E (2002). Meat consumption and colorectal cancer risk: dose-response meta-analysis of epidemiological studies. Int J Cancer. 98(2):241–56. doi:10.1002/ijc.10126 PMID:11857415

NTP (2018). Report on Carcinogens Protocol: night shift work and light at night (LAN): human cancer studies. Office of the Report on Carcinogens, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, United States Department of Health and Human Services. Available from: https://ntp.niehs.nih.gov/ntp/roc/protocols/electric_light_508.pdf.

Ross RK, Breskin A, Breger TL, Westreich D (2022). Reflection on modern methods: combining weights for confounding and missing data. Int J Epidemiol. 51(2):679–84. doi:10.1093/ije/dyab205 PMID:34536004

Smith LH, Mathur MB, VanderWeele TJ (2021). Multiple-bias sensitivity analysis using bounds. Epidemiology. 32(5):625–34. doi:10.1097/EDE.0000000000001380 PMID:34224471

Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 355:i4919. doi:10.1136/bmj.i4919 PMID:27733354

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

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

7.1 Introduction 176
7.2 Reporting considerations to aid graphical approaches to identify biases 177
7.3 Confounding 179
7.4 Information bias due to exposure and outcome misclassification 184
7.5 Selection bias 189
7.6 Conclusions 199

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.

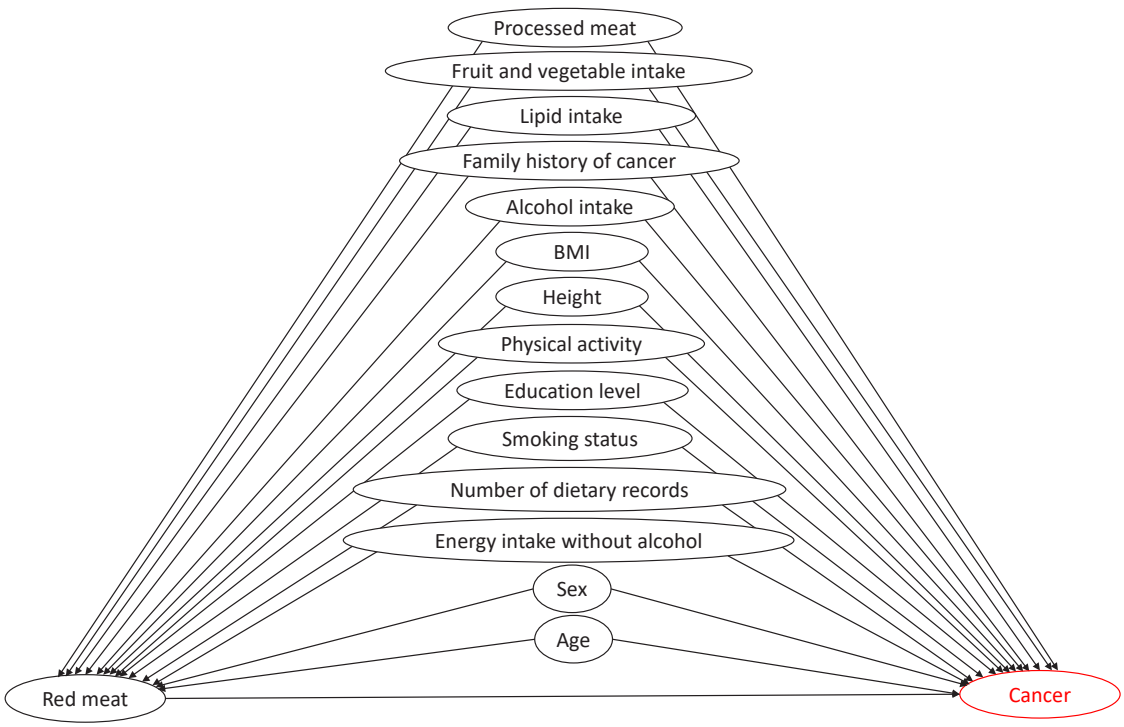
Chapter 3 discusses how to evaluate the adequacy of control for confounding in observational studies of cancer risk. This section focuses on controlling for confounding in secondary data analyses, i.e. when analysing data from case–control or cohort studies that have already been designed and conducted, 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.

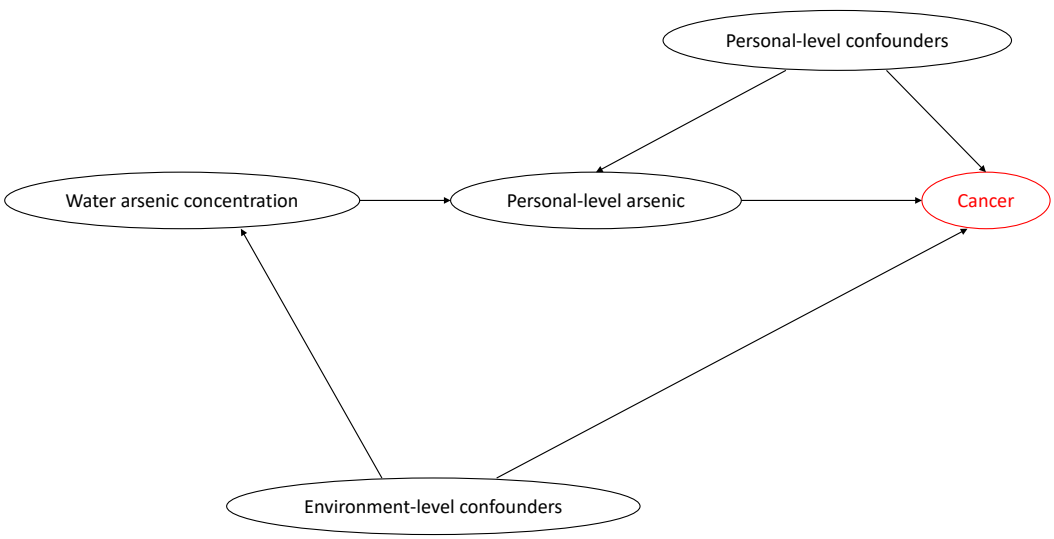


Clearly, many of the variables shown in Fig. 7.1 are likely to be important confounders (e.g. family history of cancer, smoking status). However, some of the variables included as confounders might be debatable (e.g. height), and an alternative adjustment set could be examined (e.g. by excluding height 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)

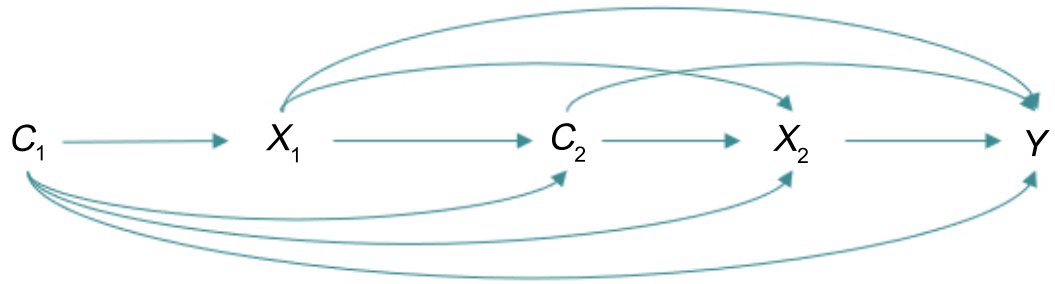
Table 7.2. Reporting considerations to facilitate bias assessment methods outlined in Section 4.3

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).

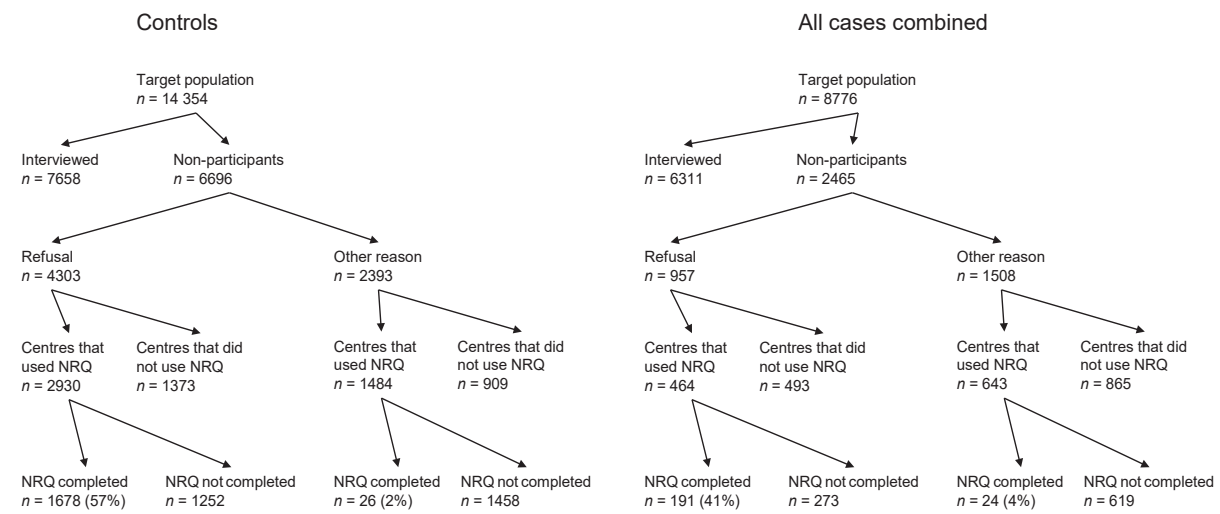
Example 7.12. Interphone study on mobile phone use and the risk of brain tumours

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

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

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

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

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

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

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

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

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/.

Annex 1. Evolution of the *IARC Monographs* Preamble from early investigations and reviews in the 1960s until the present day

A1.1 The beginnings: cancer in occupational groups 204
A1.2 Tobacco smoking and the emergence of new epidemiological methods 204
A1.3 1972: the first *IARC Monographs* 205
A1.4 1972–1980: *IARC Monographs* Volume 17 and Supplement 1 206
A1.5 1981–1990: *IARC Monographs* Supplements 4 and 7 206
A1.6 1991–2010 207
A1.7 2011 until today 207

ANNEX 1.

Evolution of the *IARC Monographs* Preamble from early investigations and reviews in the 1960s until the present day

Rodolfo Saracci and Mary K. Schubauer-Berigan

A1.1 The beginnings: cancer in occupational groups

Observations in humans pointing to life circumstances linked to the appearance of tumours go far back in history. Significant examples based on accurate observation of special population groups, rather than isolated clinical cases, have been quoted (Clemmesen, 1965): the reporting in the 16th century of a frequent respiratory disease, later identified as cancer in 1879 by Härting and Hesse, among miners in the Erzgebirge (Ore Mountains) of central Europe; the description of scrotal cancer in chimney sweepers by Pott in 1775; and the statistical evidence of an increased frequency of breast cancer in nuns presented by Rigoni-Stern in 1844, with Ramazzini's observations predating this by nearly 150 years (Franco and Franco, 2001).

However, it is since the burgeoning industrialization of the 18th century that humans have come into contact with a constantly expanding number of artificial and synthetic substances, i.e. natural substances that have been highly transformed and mixed. Specific industries or sections within industries came to represent nearly experimental situations of often prolonged and high-concentration exposure of workers to a variety of chemicals and chemical mixtures. Wilhelm Hueper, the first director of the Environmental Cancer Section at the United States National Cancer Institute, collected in a massive textbook, *Occupational Tumors and Allied Diseases* (Hueper, 1942), the accumulated evidence in humans and, through experiments in animals, on occupational exposures as causes of cancers. The documentation on cases in humans was often based on fragmentary and incomplete clinical and pathological data, and Hueper himself, not to mention his numerous critics (Sellers, 1997), regarded it as mostly circumstantial evidence of carcinogenicity, which, however, in favourable situations could justify medicolegal recognition of an occupational cause of a cancer (throughout his professional life, Hueper was a strong advocate of workers' health protection). The ultimate proof of occupational etiology of a chemical agent had to come through successful reproduction of the neoplasms in animals.

A1.2 Tobacco smoking and the emergence of new epidemiological methods

The criterion of reproduction in animals, which was in itself problematic, later proved to be a hurdle in

identifying as a cancer hazard the exposure to tobacco smoking, which after centuries of use in various forms had become widespread with the industrial production of cigarettes in the first half of the 20th century. During the same period, mortality and morbidity statistics, as well as clinical reports in several countries, indicated a marked increase of several cancers, especially of the respiratory tract, among men, suggesting a link to the spreading habit, also among men, of regular cigarette smoking. To probe this hypothesis, several studies were conducted, particularly in Germany in the years between the two World Wars (Davey Smith and Egger, 2005). A remarkable short paper by Pearl (1938) clearly showed a sizeable curtailment of the life expectancy of smokers compared with that of non-smokers.

The investigation of carcinogenicity in humans of occupational and environmental exposures and of tobacco smoking gained a renewed impetus after the Second World War. In 1950, three well-conducted case–control studies on lung cancer and cigarette smoking were published (Doll and Hill, 1950; Levin et al., 1950; Wynder and Graham, 1950); studies of worker populations accrued in the following years (Case et al., 1954; Doll, 1955). Later, the first results from cohort investigations of smoking were published (Doll and Hill, 1956; Hammond and Horn, 1958). A range of methodological issues emerged, which were unclear or even poorly understood at the time, prompting the fast development of new conceptual insights and methods of epidemiological study planning and analysis. The contributions of Cornfield are still particularly remarkable: as early as 1951, he had pointed out the essential link to risk as estimable from both cohort and case–control studies (Cornfield, 1951); in 1959, he provided a decomposition of crude risk into a net (adjusted) risk component and a component ascribable to confounding variables (Cornfield et al., 1959); and in 1962, he first used logistic regression (via discriminant analysis) to relate a dependent variable to several independent variables (Cornfield, 1962).

The time was soon ripe for two landmark publications in epidemiology: *Smoking and Health*, commissioned by the United States Surgeon General (U.S. Public Health Service, 1964), which in its conclusions indicted cigarette smoking as a cause of lung and laryngeal cancer and pipe smoking as a cause of oral cancer, and Hill's paper *The environment and disease: association or causation?* (Hill, 1965). Both publications addressed thorny issues on, and provided guidelines for, the establishment of the causal role of an exposure solely on the basis of observational studies in humans in the absence of both randomized studies in humans and reproduction of carcinogenesis in animals. The latter was the missing piece in the evidence linking tobacco smoking to cancer; both the United States Surgeon General's report and Hill's paper downplayed its role relative to epidemiological evidence, which was regarded as potentially capable of standing on its own feet. This represented a significant departure, which was bound to influence epidemiological thinking for several decades, from Hueper's criterion of reproducibility in animals, which, in turn, reflected the time-honoured etiological criteria in bacteriology (called Koch's postulates), the field of medicine in which most disease causes known at that time had been successfully identified.

Against this backdrop, two publications stand out that summarized the epidemiological evidence on cancer hazards existing by the mid-1960s: the scholarly *Statistical Studies in the Aetiology of Malignant Neoplasms* (Clemmesen, 1965) and the narrative critical review *The Prevention of Cancer: Pointers from Epidemiology* (Doll, 1967).

A1.3 1972: the first *IARC Monographs*

IARC started operating in Lyon, France, in 1967. Soon, requests were received from different public health quarters to provide an authoritative list of carcinogens for humans. Lorenzo Tomatis, who was at that time the head of the Unit of Chemical Carcinogenesis at IARC, realized that no such list could be provided without the ad hoc systematic work of assembling and evaluating all available evidence for carcinogenicity of an agent, integrating results from studies in humans and in experimental animals. The *IARC Monographs* programme was born, with the title of *IARC Monographs on the Evaluation of Carcinogenic Risk of Chemicals to Man* ("man" became "humans" in 1978), and the first volume was published in 1972. The title specified "chemicals" because this was the class of agents within which the largest number of exposures suspected to be cancer hazards were found at that time.

The first volume of the *IARC Monographs* (IARC, 1972) presented the evaluation, by a Working Group

composed of 12 scientists external to IARC, of 19 chemicals in the categories of inorganic substances, chlorinated hydrocarbons, aromatic amines, *N*-nitroso compounds, and natural products. The Working Group had met for 5 days (later to become 8 days for most *IARC Monographs* meetings) with the support of a Secretariat of IARC staff members; also attending were technical advisors, Observers, and WHO Representatives. The consensus-making body for the evaluations comprised only the Working Group members. An opening note to the reader stressed that no guiding principles were generally accepted to extrapolate results in experimental animals to humans when no data in humans were available; such principles might be developed only on the basis of some definite cases, and hence the *IARC Monographs* would continue in the initial format until sufficient background material had been accumulated. More generally, the same applied to the integration of results from human and animal studies, which for the time being could only be summarized separately, and to defining principles to weigh the evidence on carcinogenicity. For instance, the human evidence for lead and lead salts read, "There is no evidence to suggest that exposure to lead salts causes cancer of any site in man", but there was no indication of how this conclusion was reached by the Working Group. During the next 5 years, the introductory section of each *Monographs* volume was enriched by an increasingly detailed description of key points to be considered by the Working Group in reviewing and assessing the evidence. In addition to data on the chemical and physical characteristics of an agent, its uses and occurrence in the human environment, and results from cancer studies in humans and animals, other relevant biological data, in particular on mutagenicity and genotoxicity, came to be included.

A1.4 1972–1980: *IARC Monographs* Volume 17 and Supplement 1

Volume 17 of the *IARC Monographs* (IARC, 1978) had two features arising from the first years of experience. First, all introductory remarks were grouped into a Preamble, which described the *IARC Monographs* methodology and the Working Groups' operational procedures. Second, the predefined terms *sufficient evidence of carcinogenicity* and *limited evidence of carcinogenicity* were adopted, separately for animals and humans, accompanied by an outline of what types of result would support each definition.

A major advance in the evolution of the *IARC Monographs* followed 2 years later, with a Supplement to the series (IARC, 1979); a special Working Group provided some guidance for rating the evidence, separately, for studies in animals and in humans. For the latter, *sufficient* evidence indicated a causal association, *limited* evidence suggested a possible effect but was not sufficient to demonstrate a causal association, and *inadequate* evidence was considered to be qualitatively or quantitatively insufficient to permit any conclusions. As a final evaluation step, on the basis of the combined evidence from studies in animals and in humans, an agent was to be classified in one of three groups.

- Group 1: the agent is *carcinogenic to humans*. This classification was to be applied only if there was *sufficient* evidence for cancer in humans.
- Group 2, subdivided into two subcategories: Group 2A, the agent is *probably carcinogenic to humans*; Group 2B, the agent is *possibly carcinogenic to humans*. These subcategories indicate different degrees of confidence in judging the evidence as supportive of carcinogenicity.
- Group 3: the agent *cannot be classified as to its carcinogenicity to humans*.

With the introduction of this overall classification, the basic layout of the *IARC Monographs* evaluation was established. It is still maintained (see Section 1.1) as a framework suitable for incorporating updates as required by advances in cancer research.

A1.5 1981–1990: *IARC Monographs* Supplements 4 and 7

Two important steps in the evolution of the *IARC Monographs* took place in 1982 (IARC, 1982) and 1987, leading to a Preamble structure and contents that proved subject only to marginal additions for several decades. In the formulation of Supplement 7 (IARC, 1987), several types of study were enlisted to investigate cancer hazards in humans: case reports, descriptive studies of cancer occurrence in populations, and analytical case–control and cohort studies (possible intervention studies also fall into this category). On the basis of a review of findings from such studies, the evidence of carcinogenicity could be placed into one of three categories. A declaration of *sufficient* evidence of carcinogenicity indicates: "The

Working Group considers that a causal relation has been established between exposure to the agent and human cancer. That is, a positive relation has been observed between exposure to the agent and cancer in studies in which chance, bias, and confounding could be ruled out with reasonable confidence." Without this reasonable confidence, the evidence is to be rated as *limited*. If the studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of a causal association, the evidence is to be rated as *inadequate*. A fourth category, *evidence suggesting lack of carcinogenicity* (*ESLC*), as derivable from negative studies, was included with a concluding remark: "the possibility of a very small risk at the levels of exposure studied can never be excluded."

It is not coincidental that the clear and concise formulation of the criteria concerning the evidence in humans came in the years when epidemiological methods and statistical methods for epidemiology underwent in-depth revision and innovative expansion. At IARC itself, Breslow and Day began in 1976 to prepare two volumes in the Statistical Methods in Cancer Research series – to which this volume belongs – devoted, respectively, to the analysis of case–control studies (Breslow and Day, 1980) and cohort studies (Breslow and Day, 1987). As Breslow and Day (1980) stated, "The theme is, above all, one of unity. While much of the recent literature has focused on the contrast between cohort and case–control approaches to epidemiological research, we emphasize that they in fact share a common conceptual foundation, so that, in consequence, the statistical methodology appropriate to one can be carried over to the other with little or no change." The books, extensively illustrated by actual analyses of data sets from epidemiological studies, offered the best presentation, at once theoretically rigorous and practically applicable, of statistical methods in epidemiology available at the time. They became a popular reference for epidemiologists well outside the cancer field.

A1.6 1991–2010

Until 1992, the classification of an agent in Group 1 (*carcinogenic to humans*) had been strictly dependent on the existence of *sufficient* evidence from studies of cancer in humans. In 1991, in view of the continuously accruing knowledge of a variety of carcinogenesis mechanisms, a Working Group introduced a critical addition. As recorded in *IARC Monographs* Volume 54 (IARC, 1992), this reads: "Exceptionally, an agent (mixture) may be placed in this category when evidence in humans is less than sufficient but there is *sufficient* evidence of carcinogenicity in experimental animals and *strong* evidence in exposed humans that the agent (mixture) acts through a relevant mechanism of carcinogenicity."

In subsequent years, the *IARC Monographs* included three important new features. First, Volume 88 of the *IARC Monographs* (IARC, 2006) carried for the first time, in an introductory note to the reader, a much-needed terminology clarification: "The term 'carcinogenic risk' in the *IARC Monographs* series is taken to mean that an agent is capable of causing cancer under some circumstances. The *IARC Monographs* evaluate cancer hazards, despite the historical presence of the word 'risks' in the title."

Second, emphasis had constantly been placed by IARC not only on the methodological procedures used to evaluate carcinogenicity to humans but also on the objective conditions within which such evaluations were to take place. The Preamble to *IARC Monographs* Volume 94 (IARC, 2010), stemming from a review by an ad hoc advisory group, codifies in a detailed description, aimed at preventing conflicts of interest, the role of each of the five different components of participants in a *Monographs* meeting: voting Working Group members, non-voting Invited Specialists, Representatives (of national and international health agencies), scientific Observers, and the IARC staff Secretariat.

Third, in 2008 and 2009, a massive review of human carcinogens was undertaken for Volume 100 (IARC, 2012a, b, c, d, e, f), in which the data on all the agents previously classified in Group 1 (*carcinogenic to humans*) were updated and the evaluations reviewed, adding specifications of target organs. On the basis of the newly accumulated evidence, only one of the agents (human papillomavirus type 66) was moved downwards from Group 1 by the six Working Groups conducting the review.

A1.7 2011 until today

It was already apparent in the Volume 100 review (IARC, 2012a, b, c, d, e, f) that mechanistic and other relevant biological data had a steadily growing role in carcinogenicity evaluation. This promoted an overall revision of the Preamble, in 2019

(IARC, 2019a, b; Samet et al., 2020), alongside a transformation of the title to *IARC Monographs on the Identification of Carcinogenic Hazards to Humans*, which clearly defines in today's accepted terminology the programme's activity as actually implemented since the very beginning. The revision of the Preamble took into account advances in the assessment of mechanistic data, including, in particular, the identification of key characteristics of carcinogens, which provide a framework for the organization of mechanistic data and the assessment of strengths as well as gaps in evidence. The current Preamble reflects these advances and describes a process to reach a carcinogenicity classification by integrating, along parallel and harmonized lines, the three streams of evidence: experimental animal bioassays, mechanistic investigations, and epidemiological studies.

References

Breslow NE, Day NE (1980). Statistical methods in cancer research, Volume I. The analysis of case–control studies (IARC Scientific Publications No. 32). Lyon, France: International Agency for Research on Cancer. Available from: https://publications.iarc.who.int/175.

Breslow NE, Day NE (1987). Statistical methods in cancer research, Volume II. The design and analysis of cohort studies (IARC Scientific Publications No. 82). Lyon, France: International Agency for Research on Cancer. Available from: https://publications.iarc.who.int/225.

Case RA, Hosker ME, McDonald DB, Pearson JT (1954). Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. I. The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine. Br J Ind Med. 11(2):75–104. doi:10.1136/oem.11.2.75 PMID:13149741

Clemmesen J (1965). Statistical studies in the aetiology of malignant neoplasms, Volume 1. Review and results. Copenhagen, Denmark: Munksgaard Publishers.

Cornfield J (1951). A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. J Natl Cancer Inst. 11(6):1269–75. PMID:14861651

Cornfield J (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. Fed Proc. 21(4):58–61. PMID:13881407

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst. 22(1):173–203. PMID:13621204

Davey Smith G, Egger M (2005). The first reports on smoking and lung cancer: why are they consistently ignored? Bull World Health Organ. 83(10):799–800. PMID:16283059

Doll R (1955). Mortality from lung cancer in asbestos workers. Br J Ind Med. 12(2):81–6. doi:10.1136/oem.12.2.81 PMID:14363586

Doll R (1967). The prevention of cancer: pointers from epidemiology. London, UK: The Nuffield Provincial Hospitals Trust.

Doll R, Hill AB (1950). Smoking and carcinoma of the lung; preliminary report. BMJ. 2(4682):739–48. doi:10.1136/bmj.2.4682.739 PMID:14772469

Doll R, Hill AB (1956). Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. BMJ. 2(5001):1071–81. doi:10.1136/bmj.2.5001.1071 PMID:13364389

Franco G, Franco F (2001). Bernardino Ramazzini: the father of occupational medicine. Am J Public Health. 91(9):1382. doi:10.2105/AJPH.91.9.1382 PMID:11527763

Hammond EC, Horn D (1958). Smoking and death rates. Report of 44 months of follow-up of 187,783 men. J Am Med Assoc. 166(10):1159–72. doi:10.1001/jama.1958.02990100047009 PMID:13513334

Hueper WC (1942). Occupational tumors and allied diseases. Springfield (IL), USA: Charles C. Thomas.

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300. doi:10.1177/003591576506800503 PMID:14283879

IARC (1972). Some inorganic substances, chlorinated hydrocarbons, aromatic amines, *N*-nitroso compounds and natural products. IARC Monogr Eval Carcinog Risk Chem Man. 1:1–184. Available from: https://publications.iarc.who.int/19.

IARC (1978). Some *N*-nitroso compounds. IARC Monogr Eval Carcinog Risk Chem Man. 17:1–365. Available from: https://publications.iarc.who.int/35 PMID:150392

IARC (1979). Chemicals and industrial processes associated with cancer in humans. IARC Monographs, volumes 1 to 20. IARC Monogr Eval Carcinog Risk Chem Hum Suppl. 1:1–71. Available from: https://publications.iarc.who.int/133 PMID:296141

IARC (1982). Chemicals, industrial processes and industries associated with cancer in humans (IARC Monographs, volumes 1 to 29). IARC Monogr Eval Carcinog Risk Chem Hum Suppl. 4:1–292. Available from: https://publications.iarc.who.int/136 PMID:6963265

IARC (1987). Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr Eval Carcinog Risks Hum Suppl. 7:1–440. Available from: https://publications.iarc.who.int/139 PMID:3482203

IARC (1992). Occupational exposures to mists and vapours from strong inorganic acids; and other industrial chemicals. IARC Monogr Eval Carcinog Risks Hum. 54:1–336. Available from: https://publications.iarc.who.int/72 PMID:1345371

IARC (2006). Formaldehyde, 2-butoxyethanol and 1-*tert*-butoxypropan-2-ol. IARC Monogr Eval Carcinog Risks Hum. 88:1–478. Available from: https://publications.iarc.who.int/106 PMID:17366697

IARC (2010). Ingested nitrate and nitrite, and cyanobacterial peptide toxins. IARC Monogr Eval Carcinog Risks Hum. 94:1–450. Available from: https://publications.iarc.who.int/112 PMID: 21141240

IARC (2012a). Pharmaceuticals. IARC Monogr Eval Carcinog Risks Hum. 100A:1–435. Available from: https://publications.iarc.who.int/118 PMID:23189749

IARC (2012b). Biological agents. IARC Monogr Eval Carcinog Risks Hum. 100B:1–475. Available from: https://publications.iarc.who.int/119 PMID:23189750

IARC (2012c). Arsenic, metals, fibres, and dusts. IARC Monogr Eval Carcinog Risks Hum. 100C:1–501. Available from: https://publications.iarc.who.int/120 PMID:23189751

IARC (2012d). Radiation. IARC Monogr Eval Carcinog Risks Hum. 100D:1–341. Available from:https://publications.iarc.who.int/121 PMID: 23189752

IARC (2012e). Personal habits and indoor combustions. IARC Monogr Eval Carcinog Risks Hum. 100E:1–575. Available from: https://publications.iarc.who.int/122 PMID:23193840

IARC (2012f). Chemical agents and related occupations. IARC Monogr Eval Carcinog Risks Hum. 100F:1–599. Available from: https://publications.iarc.who.int/123 PMID:23189753

IARC (2019a). Report of the Advisory Group to Recommend an Update to the Preamble to the *IARC Monographs*. Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/wp-content/uploads/2019/09/Preamble-AG-Report_final-web.pdf.

IARC (2019b). Preamble to the *IARC Monographs* (amended January 2019). Lyon, France: International Agency for Research on Cancer. Available from: https://monographs.iarc.who.int/iarc-monographs-preamble-preamble-to-the-iarc-monographs/.

Levin ML, Goldstein H, Gerhardt PR (1950). Cancer and tobacco smoking; a preliminary report. J Am Med Assoc. 143(4):336–8. doi:10.1001/jama.1950.02910390008002 PMID: 15415261

Pearl R (1938). Tobacco smoking and longevity. Science. 87(2253):216–7. doi:10.1126/science.87.2253.216 PMID:17813231

Samet JM, Chiu WA, Cogliano V, Jinot J, Kriebel D, Lunn RM, et al. (2020). The *IARC Monographs*: updated procedures for modern and transparent evidence synthesis in cancer hazard identification. J Natl Cancer Inst. 112(1):30–7. doi:10.1093/jnci/djz169 PMID: 31498409

Sellers C (1997). Discovering environmental cancer: Wilhelm Hueper, post-World War II epidemiology, and the vanishing clinician's eye. Am J Public Health. 87(11):1824–35. doi:10.2105/AJPH.87.11.1824 PMID:9366640

U.S. Public Health Service (1964). Smoking and health: report of the Advisory Committee to the Surgeon General of the Public Health Service (Public Health Service Publication No. 1103). Washington (DC), USA: U.S. Department of Health, Education, and Welfare, Public Health Service.

Wynder EL, Graham EA (1950). Tobacco smoking as a possible etiological factor in bronchiogenic carcinoma: a study of six hundred and eighty-four proved cases. JAMA. 143(4):329–36. doi:10.1001/jama.1950.02910390001001 PMID:15415260

ANNEX 2.

Examples for which worked spreadsheets or R code are provided

This annex provides a list of the examples for which worked spreadsheets or R code are provided. These Supplementary Materials are available online from https://publications.iarc.who.int/634#supmat.

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

Example 4.16. Analysis of bias from non-differential exposure misclassification 101

Example 4.17. Analysis of bias from differential exposure misclassification 102

Example 4.21. Probabilistic bias analysis for exposure misclassification 106

Example 4.25. Opium use and HNSCC – bias analysis for categorical data.....114

Chapter 5. Selection bias and other miscellaneous biases

Example 5.30b. Quantitative bias analysis to examine potential selection bias in a nested case–control study of breast cancer 147

Example 5.30e. Applying probabilistic bias analysis results to estimated odds ratios in a nested case–control study of breast cancer 149

Chapter 6 and Annex 3. Incorporating bias assessments into evidence synthesis

Example 6.9. Multiple bias analysis 172

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

Example 7.14. Selection probabilities in the Interphone study 195

Example 7.16. Case–control study of opium exposure and oesophageal cancer 199

Annex 3. Incorporating bias assessments into evidence synthesis

A3.1 Multiple-bias analysis: worked examples 214
A3.2 Overview of multiple-bias analysis of the data of Aliramaji et al. (2015) using pseudo-data and bias-factor approaches 217
A3.3 Sensitivity analysis 226
A3.4 A potential probabilistic multiple-bias analysis strategy 226

ANNEX 3.

Incorporating bias assessments into evidence synthesis

Alexander P. Keil

A3.1 Multiple-bias analysis: worked examples

The goal of bias analysis is to estimate the expected effect that would have been estimated in the study, had that study not been subject to the bias of concern. For the purposes of this book, estimates of the risk ratio are considered, contrasting the cumulative risk of an outcome at two levels of exposure. The risk ratio obtained in a given study is referred to as the apparent risk ratio, RR_{app}, and the risk ratio after performing bias analysis is referred to as the adjusted risk ratio, RR_{adj}. The true target of RR_{adj} is $RR_{unbiased}$, the risk ratio that would have been estimated in the absence of any systematic bias, but RR_{adj} is used to emphasize the necessary simplifying assumptions that feed into a bias analysis and the reality that the only bias parameters that may be available are typically, at best, approximations to the true bias parameters.

The following worked example offers a template for adapting multiple-bias analysis to new scenarios, but it also indicates a unique aspect of multiple-bias analysis that sets it apart from single-bias analyses: the approach to serial multiple-bias analysis ought to vary according to the order in which biases are thought to occur for the RR_{app} under consideration. Smith et al. (2021, p. 627) write, "In general, we can think of biases as layers that we must peel off sequentially and the order in which we do so is the reverse of the order in which they occurred in the data." Fox et al. (2021, pp. 416–417) state, "Bias-adjustment does not generally reduce to independent multiplicative bias factors […], so the order of bias-adjustments can affect the ultimate result."

Two primary approaches to bias analyses could be considered: (i) the approach of Smith et al. (2021), which uses (dependent) bias factors to estimate upper or lower bounds of bias for a range of bias parameters, and (ii) an approach given in Fox et al. (2021) that involves the calculation of pseudo-data. In the first approach, a bound of the value of RR_{adj} is established that is typically a direct answer to the question "What is the most extreme value of the true risk ratio that is still consistent with RR_{app} under the bias parameters?" This approach answers the useful question (for hazard identification) "Can we rule out bias as the sole explanation of a non-null effect estimate?" However, the approach described by Smith et al. (2021) has not been studied extensively, so it is not known how conservative the bound is (e.g. how likely it is that a bound will indicate consistency with

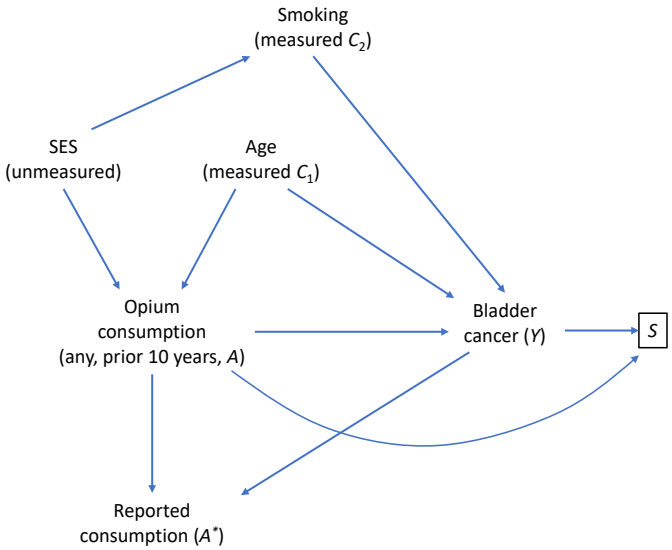
the null hypothesis). Furthermore, this bounding approach does not easily accommodate scenarios in which biases may offset each other. The second approach establishes a value for RR_{adj} that is derived by stacking the approaches used in other chapters of this book for individual biases. Rather than a bound, the second approach provides a single best estimate under a given set of bias parameters. This approach can be extended in a probabilistic bias analysis to accommodate uncertainty in the bias parameters.

A3.1.1 Worked example with a single study on opium consumption and bladder cancer

The primary examples are based on a study on opium consumption and bladder cancer (Aliramaji et al., 2015). Fig. A3.1 expresses three potential biases that might be considered. Specifically, these are related to issues of exposure misclassification, unmeasured confounding, and selection bias, the last of which arises from the method of selecting the study population. In the analysis of Aliramaji et al. (2015), which is a case–control study, odds ratios, rather than risk ratios, are reported. The bias correction methods used here rely variously on odds ratios and risk ratios, but note that, given a rare disease like bladder cancer, these can be considered nearly equivalent so that methods to adjust a risk ratio can be used with an odds ratio.

Aliramaji et al. (2015) do not report measures of association, but the apparent odds ratio is calculable as the crude odds ratio (Table A3.1). Note that the crude odds ratio of 2.72 is different from the odds ratio given in Table 6.4, which was calculated as the crude odds ratio among tobacco non-smokers. The odds of having bladder cancer (case odds) are 2.15 among those who reported opium use and 0.79 among those who did not, leading to an apparent odds ratio of 2.72. As noted, for this odds ratio there is a concern over unmeasured confounding (because an adjusted estimate of association for key confounders of age and smoking was not reported). Selection bias concerns arise because the control participants were selected from among hospitalized patients who were being treated surgically for gall bladder stones; in addition, as noted in Fig. A3.2, selection bias may arise because there was frequency matching on sex (Mansournia et al., 2018). Opium exposure misclassification concerns arise because these were identified in prior validation studies on self-reported opium consumption. This bias analysis is focused on the scenario in Fig. A3.2, in which selection bias due to matching is of greater concern than Berkson bias (which was discussed in Chapter 5), because of the recruitment of hospitalized control participants.

Fig. A3.1. Simplified directed acyclic graph (DAG) showing potential areas where bias correction may be used in an analysis of a study on opium consumption and bladder cancer (Aliramaji et al., 2015). This DAG illustrates three biases: (i) differential exposure measurement error: measured exposure (A^*) is a mismeasured proxy of true exposure (A), in which measurement error depends on the outcome (Y); (ii) selection bias: the recruitment of hospitalized control participants raises concerns that selection in the study may be affected by opium use, because opium use can cause other hospitalizable outcomes; and (iii) confounding by age and smoking (C_1, C_2). SES, socioeconomic status.



A3.1.2 Order of bias corrections

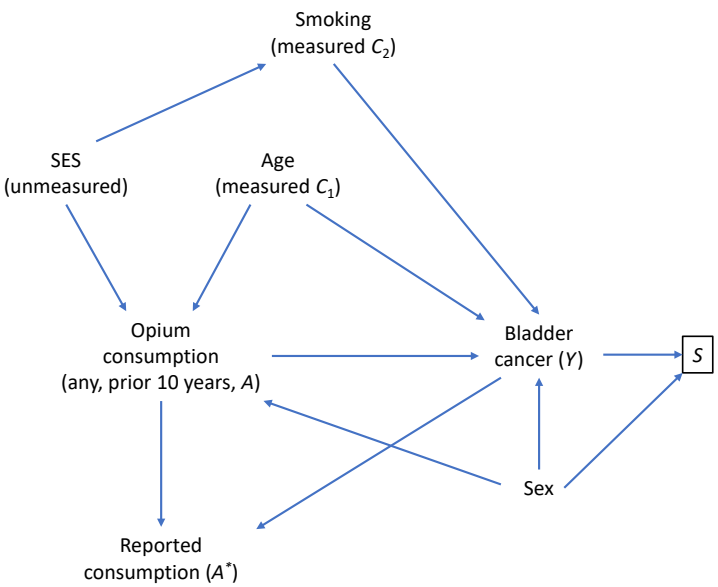
This example, drawn from Aliramaji et al. (2015), is an interesting case study in multiple-bias analysis because it demonstrates issues of measured

Table A3.1. Raw exposure and case status data, and calculated crude odds ratio

		Case participants	Control participants	Case odds
Opium use	Yes	58	27	2.15
	No	117	148	0.79
Crude/apparent odds ratio				2.72

Source: Aliramaji et al. (2015).

Fig. A3.2. Simplified directed acyclic graph (DAG) showing alternative specification in which bias correction may be used in an analysis of a study on opium consumption and bladder cancer (Aliramaji et al., 2015). This DAG illustrates three biases: (i) differential exposure measurement error: measured exposure (A^*) is a mismeasured proxy of true exposure (A), in which measurement error depends on the outcome (Y); (ii) selection bias: hospitalized control participants were being treated surgically for gall bladder stones, for which there is no known association with bladder cancer; however, control participants were matched with case participants on sex; this latter factor was left uncontrolled in the analysis, leading to selection bias; (iii) confounding by age and smoking (C_1, C_2). SES, socioeconomic status.



confounding, effect measure modification, and selection bias arising from matching on confounders. The order of bias correction is guided by the order in which biases may appear in the data. In this study, frequency matching of the study design on confounding factors (age and sex) can introduce selection bias in analyses that are unadjusted for these factors. Thus, selection bias occurs because of an open path from the outcome to the (correctly classified) exposure through the selection node and the frequency-matched factors. Adjusting for this bias requires that pathway to be closed; this can be done by adjusting for the matched factors. Selection bias can thus be considered as the first bias to address, given that true exposure need not be measured. Had exposure directly influenced selection (as might occur if opium were a cause of a condition that resulted in a person's being selected as a control participant), misclassification bias would necessarily first have been considered for adjustment. Once study selection is adjusted for, exposure misclassification can be adjusted for. Finally, consideration is given to confounding, which is considered to be a function of reality, rather than study design or measurement issues. This bias would therefore be considered to happen first (in temporal order), and correction for it would come last. Smith et al. (2021) and Fox et al. (2021) both consider alternative orderings in multiple-bias analysis.

A notable issue when selecting the order of bias correction for this example is in regard to the available data. Fox et al. (2021, p. 417) perfectly encapsulate this scenario: "Classification parameters might be measured in a population-based setting (i.e. negligible selection bias), but be applied to a data set where selection bias is a concern. In this setting, the analyst should bias-adjust for selection bias before bias-adjusting for misclassification, even if the selection bias preceded the misclassification in the data generation process." Thus, even if true exposure plays a role in selection bias, the role of opium in inducing selection bias is less pertinent to the biases at issue than the role of misclassification in effect estimation. Thus, there is an additional reason in this example to apply misclassification parameters in data that are already adjusted for selection bias, rather than the other way around.

A3.2 Overview of multiple-bias analysis of the data of Aliramaji et al. (2015) using pseudo-data and bias-factor approaches

The pseudo-data approach (using bias parameters to calculate bias-adjusted data, which can then be analysed as though they were real data) is used for plausible (under reasonable assumptions) ranges of bias arising from exposure misclassification and selection bias. The pseudo-data approach begins with an apparent odds ratio that is unadjusted for confounders. Consideration is also given to measured confounding, similar to the selection bias adjustment example of Chapter 4. The bias-factor approach, as described by Smith et al. (2021), is used to place a lower bound on the multiple-bias-adjusted odds ratio, given a set of bias parameters. The bias-factor approach begins with an apparent odds ratio that is adjusted for confounders. The parameters needed in the bias-factor approach are summarized in Table A3.2. Both approaches are demonstrated in the R code provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat).

A3.2.1 The apparent odds ratio as the basis for subsequent calculations

The crude odds ratio of Table A3.1 can be used as the apparent odds ratio for subsequent analysis in the pseudo-data approach. Alternatively, Aliramaji et al. (2015) give enough information to infer smoking-stratified results (in the results, sample sizes were reported for case and control participants who had both consumed opium and smoked cigarettes for longer than 1 year); these figures are given in Table A3.3 and yield a smoking-adjusted odds ratio of 1.25 from unconditional logistic regression and stratum-specific odds ratios of 4.1 among non-smokers and 0.5 among smokers. The stratum-specific results indicate substantial odds ratio modification; this is a key consideration, as discussed for unmeasured confounding in Section 3.3. This adjusted odds ratio can, nonetheless, be selected as the apparent odds ratio for the bias-factor approach because it does not rely on tabulated data.

A3.2.2 Selection bias adjustment using bias factors

As shown in Table A3.2, adjustment for selection bias using the bias-factor approach involves consideration of a factor, U_S, that influences selection into the study and is also conditionally associated with the outcome. In this analysis, U_S is considered as sex only, given that opium use varies strongly with sex, and sex was used as a matching factor for the study but was not subsequently adjusted for. Matching without adjustment for sex created a backdoor biasing pathway, because of conditioning on the collider S (opium consumption ← sex → S ← bladder cancer). For a causal interpretation of the selection-bias adjustment, adjustment for U_S should be sufficient to render the study outcome and study selection (the node S in Figs. A3.1 and A3.2) independent, given other factors that are included in the analysis. Crucially, the selection bias under consideration here affects the meaning of S. By definition, $S = 1$ is the value of S for members in the study population; $S = 0$ is given for individuals who would have been part of the study data, had they not been selected out of the study. Generally, to adjust for selection bias, one must know or assume something about those for whom $S = 0$. Here, those for whom $S = 0$ are a (potentially hypothetical) group of women who were at the hospital used in the study for surgical treatment of gall bladder stones but were not included in the study. To simplify further calculations, it is assumed that selection into this eligible population is not related to sex; thus, a similar sex ratio among potential control participants to that in the underlying source population (i.e. 1:1) is expected.

Calculating the bias factor for this example involves specifying (for binary misclassified exposure A^*) $\Pr(U_S = u \mid A^* = a, S = a, C = c)$, $\Pr(U_S = u \mid A^* = a, S = 1 - a, C = c)$ (the prevalence parameters), and $\Pr(Y = 1 \mid A^* = a, C = c, U_S = u)$ (the risk parameters), which are the prevalence of the unmeasured factor at some level u (i.e. male or female for the binary in this example) and the risk of the outcome at specified values of A^*, C, and U_S.

To inform the prevalence parameters, data were included from a national survey of residents of the Islamic Republic of Iran conducted by Moradinazar et al. (2020), who estimated the prevalence of drug use, stratified by several demographic variables, using the survey question "Have you used illicit drug more than one time during a lifetime?" The average prevalence was estimated as 24.1% among men and 2.2% among women. These survey data correspond to the sex-specific prevalences, $\Pr(A^* = a \mid U_S = u)$; it is assumed that this does not vary meaningfully across levels of covariates C

Table A3.2. Summary of parameters that determine bias factors according to the multiple-bias bounding method of Smith et al. (2021)

Type of bias[a]	Identifier	Parameter	Definition and notes	Investigator-specified values	Calculated value
Unmeasured confounding	D	$\Pr(U_c = u \mid A = 1, C = c)$	Prevalence of $U_c = u$, among exposed, given observed confounders.	✓	
	E	$\Pr(U_c = u \mid A = 0, C = c)$	Prevalence of $U_c = u$, among unexposed, given observed confounders.	✓	
	RR_{AU_C}	$\max_u(D/E)$	The maximum factor (over levels of the unmeasured confounder) by which exposure is conditionally associated with a given value of the unmeasured confounder in an analysis free of selection bias and misclassification bias. For binary U_c, this is the maximum prevalence ratio contrasting levels of exposure, given measured confounders.		✓
	F	$\Pr(Y = 1 \mid A = a, C = c, U_c = u)$	The risk of Y, given exposure, observed confounders, and unmeasured confounders.	✓	
	RR_{U_CY}	$\max_a[\max_u(F)/\min_u(F)]$	The maximum value (across levels of exposure) of the ratio of the maximum risk (across different levels of U_c) and minimum risk (across different levels of U_c). This is the maximum possible risk ratio contrasting the outcome risk across levels of U_c and describes confounding bias above and beyond measured confounding in an analysis free of selection bias and misclassification bias.		✓
	BF_C	$g(RR_{U_CY}, RR_{AU_C})$	Multiplicative bias factor, confounding. This is interpreted as the confounding-bias risk ratio or the ratio of the risk ratio adjustment for U_c to the risk ratio with unmeasured confounding by U_c.		
Differential or non-differential exposure misclassification	G	$\Pr(A^* = 1 \mid Y = 1, A = 0, S = 1, C = c)$	False-positive probability among case participants (1 − specificity)	✓	
	H	$\Pr(A^* = 1 \mid Y = 0, A = 0, S = 1, C = c)$	False-positive probability among non-case participants (1 − specificity)	✓	
	I	$\Pr(A^* = 1 \mid Y = 1, A = 1, S = 1, C = c)$	True-positive probability among case participants (sensitivity)	✓	
	J	$\Pr(A^* = 1 \mid Y = 0, A = 1, S = 1, C = c)$	True-positive probability among non-case participants (sensitivity)	✓	
	FPOR	$(G/H)/[(1 - G)/(1 - H)]$	False-positive odds ratio		✓
	SEOR	$(I/J)/[(1 - I)/(1 - J)]$	Sensitivity odds ratio		✓
	CCR	$(I/J)/[(1 - G)/(1 - H)]$	Correct classification ratio		✓
	ICR	$(G/H)/[(1 - I)/(1 - J)]$	Incorrect classification ratio		✓
	OR_{A^*Y}	max(FPOR, SEOR, CCR, ICR)	Maximum selection odds ratio		✓
	BF_M	OR_{A^*Y}	Multiplicative bias factor, differential exposure misclassification. Note that this is a bias odds ratio and applies when the effect estimate is an odds ratio. In rare disease settings, this approximates the risk ratio and can be used for risk ratios.		✓

Table A3.2. Summary of parameters that determine bias factors according to the multiple-bias bounding method of Smith et al. (2021) (continued)

Type of bias[a]	Identifier	Parameter	Definition and notes	Investigator-specified values	Calculated value
Selection bias	K	$\Pr(Y = 1 \mid A = a, C = c, U_S = u)$	The risk of Y, given exposure, observed confounders, and an unmeasured source of selection bias	✓	
	$RR_{U_S Y}(a)$	$\max(K)/\min(K)$	The ratio (at a given level of exposure, and given confounders) of the maximum risk of the outcome (across levels of the unmeasured source of selection bias) and the minimum risk of the outcome (across levels of the unmeasured source of selection bias). This is the maximum possible risk ratio contrasting levels of the variable that is a source of selection bias. If U_S is binary, this is simply the conditional risk ratio contrasting $U_S = 1$ against $U_S = 0$, or its inverse, whichever is larger.		✓
	L	$\Pr(U_S = u \mid A = a, S = a, C = c)$	Prevalence of $U_S = u$ at a given level of exposure, among those who were selected into the study (if exposed) or among those not selected into the study (if unexposed), given observed confounders	✓	
	M	$\Pr(U_S = u \mid A = a, S = 1 - a, C = c)$	Prevalence of $U_S = u$ at a given level of exposure, among those who were selected into the study (if considering the unexposed) or among those not selected into the study (if considering the exposed), given observed confounders	✓	
	$RR_{SU_S}(a)$	$\max(L/M)$	The maximum ratio by which selection into the study increases the prevalence of some value of U_S, within strata of exposure, given confounders. For binary U_S, this is the prevalence ratio for U_S, given exposure and confounders, comparing those selected into the study versus those who are not selected, or its inverse, whichever is larger.		✓
	BF_S	$g[RR_{U_S Y}(a = 1), RR_{SU_S}(a = 1)] \times g[RR_{U_S Y}(a = 0), RR_{SU_S}(a = 0)]$	Multiplicative bias factor, selection bias. This is interpreted as the selection-bias risk ratio or the ratio of the risk ratio under no selection bias to the risk ratio with selection bias.		✓

A, exposure of interest; C, measured confounders; OR, odds ratio; Pr, prevalence; RR, risk ratio; S, selection into study; U_C, unmeasured factor that introduces confounding bias (unmeasured confounder); U_S, unmeasured factor that introduces selection bias; Y, outcome of interest.

[a] The biases included in this particular bias analysis include (binary) differential exposure misclassification among the study population, selection bias in which bias can be envisioned as selection on a factor that results in the expected effect in the study population differing from the expected effect in the target source population, and unmeasured confounding that results from a single confounder. Note that this table presents one possible set of hypothesized biases; Smith et al. (2021) discuss a broader set of potential bias combinations for which multiple-bias bounding can be used. The function $g(a, b) = (a \times b)/(a + b - 1)$ is given by Smith et al. (2021).

Source: Smith et al. (2021).

Table A3.3. Tobacco smoking-stratified estimates of the odds ratio, and summary adjusted odds ratio (via logistic regression) inferred from Table A3.1 and results reported from Aliramaji et al. (2015)[a]

		Bladder cancer			
		Smokers		Non-smokers	
		Case participants	Control participants	Case participants	Control participants
Opium use	Yes	44	20	14	7
	No	50	11	67	137
	Total	94	31	81	144
Stratified odds ratio (smokers)				0.48	
Stratified odds ratio (non-smokers)				4.09	
Summary odds ratio				1.25	

[a] Numbers of case and control participants with exposure to both smoking and opium use, as well as marginal totals of smokers and non-smokers by case status, were given in the paper; this information could be used to complete the table.

(this may not hold well for tobacco smoking, which was strongly related with drug use in the survey, but there are insufficient data to proceed using a smoking-specific correction, given the limitations of performing bias corrections on published data). Under the assumption that exposure itself does not affect selection, given sex, prevalences from the survey can be expanded to $\Pr(A^* = a \mid U_S = u) = \Pr(A^* = a \mid U_S = u, S = s)$ and the sex-specific survey data can be used in further calculations. Some selection-bias adjustment parameters will also be based on study data, but note that the effect estimate in the study (adjusted odds ratio) is used to approximate the adjusted risk ratio from a cohort analysis, so selection parameters that rely on the study data should be estimated from only the data for control participants. In an unmatched case–control study, the control participants should represent the distribution of exposures in the source population. In this matched setting, the control participants represent a stratified sample from the source population, where the sampling proportions are derived from the distribution of sex among the case participants.

The sex-specific prevalences of drug use from the survey can then be transformed to yield prevalences of each sex in each category of exposure, using Bayes' theorem. This is given as:

$$\Pr(U_S = u \mid A^* = a, S = s) = \Pr(A^* = a \mid U_S = u, S = s) \times \Pr(U_S = u \mid S = s)/\Pr(A^* = a \mid S = s) \quad \text{(A3.1)}$$

The multiplicative factor $\Pr(U_S = u \mid S = s)/\Pr(A^* = a \mid S = s)$ can be estimated from study data, demographic data, and the population distribution of sex (here, a 1:1 female:male ratio is assumed). The sex-specific proportions in the control data are $\Pr(U_S = u \mid S = 1)$, and are given as 87.4% for men and 12.6% for women. The distribution of exposure in the control participants is given as $\Pr(A^* = a \mid S = 1)$ (15.4% exposed, 84.6% unexposed). The sex-standardized survey estimate of drug use prevalence (24.1% × 0.5 + 2.2% × 0.5 = 13.2%) is used as the assumed exposure prevalence in the target control population, and can be given as

$$\Pr(A^* = a) = \Pr(S = 1)\Pr(A^* = a \mid S = 1) + \Pr(S = 0)\Pr(A^* = a \mid S = 0) \quad \text{(A3.2)}$$

This enables solving for $\Pr(A^* = a \mid S = 0)$, noting that the selection probabilities are derived by dividing the number of control participants by the expected number of control participants if sex had not been used as a matching factor, (153 men + 22 women)/(153 + 153) = 57%, and $\Pr(A^* = 1 \mid S = 0) = 10.1\%$, which supports the idea that the unselected population will have less exposure than the selected control participants (prevalence = 15.4%), because the unselected population will include women who were omitted from the study as a consequence of matched sampling. However, this percentage is substantially higher than the female-specific prevalence of drug use in the survey data of 2.2%, which is an alternative value that could be used in a sensitivity analysis.

Finally, note that $\Pr(A^* = a \mid U_S = u, S = s) = \Pr(A^* = a \mid U_S = u)$, through the assumptions of Fig. A3.2, because selection and exposure are independent, given sex. Thus, exposure

prevalence by sex can be taken directly from the survey data estimates of 24.1% among men and 2.2% among women. Calculations of $\Pr(U_S = u \mid A^* = a, S = s)$ can then be made by application of Bayes' theorem, as before.

The risk parameters are used in the bias-factor method by taking the maximum ratio (at each level of exposure) by which U_S could increase the risk. Consequently, the exact risks are not crucial, but the risk ratio comparing bladder cancer across levels of U_S is crucial. For non-binary U_S, the risk ratio would be calculated for the lowest risk value of U_S against the highest risk value of U_S. Here, it is possible simply to fill in an arbitrary (valid) value for the risks for unexposed (or exposed) men and use the risk ratio for being female compared with that of being male among the unexposed (or exposed) participants to calculate the second set of risks. Ideally, these values could be informed through regression coefficients for sex from a study in which sex was included in a model for the risk of bladder cancer, given opium use. This would be different from a crude risk ratio contrasting men and women, because opium use is a potential mediator between sex and bladder cancer, and the parameter needed in this case is the risk ratio for sex with adjustment for opium use as a mediator. Such coefficients may not be available in the literature because they may not be a central feature of interest in a regression analysis. For example, Hadji et al. (2022) estimate an opium-adjusted odds ratio for sex but do not report the coefficient for sex in the model. Consequently, the crude odds ratio estimate for sex, 0.33, given by Hadji et al. (2022) is used to approximate the opium-conditional odds ratio. Negligible effect measure modification by sex occurred for opium use, suggesting that similar risk ratios can be used for men and women. After filling in arbitrary values for exposed and unexposed men of 0.08 and 0.02, respectively, and letting the odds ratio estimate for sex (0.33) stand in for the estimated risk ratio, risk estimates for exposed and unexposed women of 0.026 and 0.0066, respectively, are used. Note that the method is not sensitive to the absolute values of risk or to the ratio of risks between exposed groups.

Finally, given that the prevalence and risk parameters have been fully enumerated, the bias-factor calculation leads to a selection-bias factor of BF_S = 1.40; after adjustment for selection bias, the lower bound of the adjusted risk ratio is 1.24/1.40 = 0.89. At this point one might stop, if the goal is to determine whether the plausible lower bound moves across the null from the study estimates, because further adjustments will only decrease this bound. One might also refine selection-bias adjustment by calculating an additional value of BF_S for the impact of selection bias by recruiting hospitalized control participants or matching on age; the lower bound of 0.89 would be divided by this additional factor to obtain a new lower bound. This use of the survey data from Moradinazar et al. (2020) demonstrates that data from outside sources can inform bias analysis in useful ways, even if the bias parameters that are needed for analysis are not estimated directly in the study, provided that additional calculations can be performed, as was true here.

A3.2.3 Selection bias adjustment using pseudo-data

The selection bias induced by matching on sex and age means that in a study sample without this selection a different distribution of these two factors would be observed. A simple (and long-used) approach to estimate the effect of selection bias for an odds ratio (which is how the impact of opium use on bladder cancer was estimated by Aliramaji et al., 2015) is to multiply the odds ratio by the selection odds, which are calculated using the probability of selection into the study for the four combinations of case or control status and exposed or unexposed status. Fox et al. (2021) show that this is equivalent to inverse odds-of-selection weighting in this simple case of four selection parameters. Inverse odds-of-selection weighting could be extended further, to account for selection bias that occurs specifically as a result of matching on sex and age, but such an approach would rely on having much more refined estimates than are available in the study of Aliramaji et al. (2015). Thus, the simpler approach to weighting is chosen here; this is equivalent to the selection-odds approach.

The selection probabilities for the combination of case or control status and exposed or unexposed status can be inferred partly by the study design. Because the concern for selection bias is matching of the control participants, there is no issue (in this situation) with selection of the case participants in terms of bias. It would be expected that the proportion of exposed case participants in the study is equal to the proportion of exposed case participants in the source population, such that the selection probability of case participants

can be considered to be 1.0, i.e. there are no additional case participants who would have been observed in the population if matching had not been used (however, this would not be the situation if exact matching led to the exclusion of some unmatched case participants).

The selection proportions for the exposed and unexposed control participants can then be informed by the same survey data as before, given the rarity of bladder cancer in the population. These probabilities are given by $\Pr(S = 1 \mid Y = 0, A^* = 1)$ and $\Pr(S = 1 \mid Y = 0, A^* = 0)$, which are not directly given by the data or in validation data. However, these can be expanded to include sex by noting that

$$\Pr(S = 1 \mid Y = 0, A^* = a) = \sum_u \Pr(S = 1 \mid Y = 0, A^* = a, U_S = u) \Pr(U_S = u \mid A^* = a, Y = 0) \quad \text{(A3.3)}$$

which are the sex-specific (and exposure-specific) probabilities of selection and the population exposure-specific probabilities of reporting sex as male or female.

First, it is necessary to find $\Pr(S = 1 \mid Y = 0, A^* = a, U_S = u)$. Because selection into the study did not depend on exposure, conditional on sex, $\Pr(S = 1 \mid Y = 0, A^* = a, U_S = u)$ is equal to $\Pr(S = 1 \mid Y = 0, U_S = u)$. Next, using Bayes' theorem:

$$\Pr(S = 1 \mid Y = 0, U_S = u) = \Pr(U_S = u \mid S = 1, Y = 0)\Pr(S = 1 \mid Y = 0)/\Pr(U_S = u \mid Y = 0) \quad \text{(A3.4)}$$

The value of $\Pr(U_S = u \mid S = 1, Y = 0)$ is given by the study data as 87.4% for men and 12.6% for women. $\Pr(U_S = u \mid Y = 0)$ is assumed to be 50% (1:1 female:male ratio in the source population, and a rare outcome, such that the ratio in the non-case participants will be very similar). $\Pr(S = 1 \mid Y = 0)$ is the overall selection proportion for non-case participants, for whom direct data are not available but can be derived using the idea that a control group with no selection bias ought to have a 1:1 sex ratio, and would thus be expected to include 153 men and 153 women. This means that the probability of selection is (153 + 22)/(153 + 153) = 57.1%. As an example calculation for (exposed and unexposed) women, $\Pr(S = 1 \mid Y = 0, A^* = a, U_S = \text{female})$ is given as 0.126 × 0.571/0.5 = 0.143; for men, this is 0.874 × 0.571/0.5 = 0.998.

To complete the selection probabilities, it is also necessary to find $\Pr(U_S = u \mid A^* = a, Y = 0)$. Unfortunately, these probabilities are not given in the study data, but noting that for a rare disease $\Pr(U_S = u \mid A^* = a, Y = 0) \approx \Pr(U_S = u \mid A^* = a)$, this quantity can be estimated from survey data and (again) Bayes' theorem. First,

$$\Pr(U_S = u \mid A^* = a) = \Pr(A^* = a \mid U_S = u)\Pr(U_S = u)/\Pr(A^* = a) \quad \text{(A3.5)}$$

As before, exposure prevalence by sex can be taken directly from the survey data estimates of $\Pr(A^* = a \mid U_S = \text{male}) = 24.1\%$ and $\Pr(A^* = a \mid U_S = \text{female}) = 2.2\%$. A 1:1 sex ratio yields $\Pr(U_S = u) = 50\%$, and (for example) the marginal probability of misclassified exposure is given by the sex-standardized probability of exposure:

$$\sum_u \Pr(U_S = u)\Pr(A^* = 1 \mid U_S = \text{male}) = 0.5 \times 0.022 + 0.5 \times 0.241 = 0.1315 \quad \text{(A3.6)}$$

The full calculation yields selection probabilities of 0.93 for exposed control participants and 0.52 for unexposed control participants, which yields a selection-bias-adjusted relative risk of 4.9 (Table A3.4).

Although the selection-bias-adjusted relative risk of 4.9 obtained using the pseudo-data method and the lower bound relative risk of 0.89 obtained using the bias-factor method seem to give conflicting results, there are important caveats to note. First, the bias-factor method is focused on extreme circumstances, such that even if, in expectation, a bias might be downwards, the bias-factor method

Table A3.4. Selection-bias-adjusted pseudo-data, selection probabilities, and calculated selection-bias-adjusted risk ratio

		Bladder cancer		Case odds/OR
		Case participants	Control participants	
Opium use	Yes	58	29.08	1.99
	No	117	285.76	0.41
Crude/apparent odds ratio				4.87
Selection probabilities	Exposed	1	0.928	
	Unexposed	1	0.518	

OR, odds ratio.
Source: Aliramaji et al. (2015).

focuses on extrema of the bias, which could be in opposing directions from the expectation. Second, the process of adjusting for selection bias from matching using the pseudo-data method reintroduced confounding by sex (which was presumably what matching was intended to solve). Because being male is strongly positively associated with both opium use and bladder cancer, this induced confounding is expected to be upwards, such that a confounding-bias-adjusted relative risk would be expected to be less than 4.87. The topic of confounding by sex will be revisited in Section A3.2.6.

A3.2.4 Exposure misclassification bias adjustment using bias factors

Exposure misclassification is a special concern in studies of illicit drug use when self-report is used to determine drug use and in any study in which recall periods are long for defining exposure. In the study of Aliramaji et al. (2015), a hospital-based study in the Islamic Republic of Iran, patients were considered exposed if opium consumption was noted in their files from the pathology department, hospital archives, and phone calls (although scant details are given, each of these records presumably originates from self-report or physician report, rather than routine biological test results). Exposure was defined as reported duration of use greater than or equal to 1 year.

Correction for exposure misclassification via bias factors requires values of specific sensitivity and specificity for case and control participants. Ideally, a validation study to adjust the estimates of Aliramaji et al. (2015) would be able to provide estimates of each of these parameters. The most relevant study that could be identified was that of Rashidian et al. (2017), who conducted an assay-based validation study of self-reported opioid use (primarily raw opium) among patients in a hospital in the Islamic Republic of Iran (also the study population setting and country of origin for the analysis of Aliramaji et al., 2015). Self-reported regular use of opioids for 6 months or longer in the user's lifetime was selected as the target variable, which was validated by two measures: self-reported use in the previous 72 hours and, among those who did not self-report use in the previous 72 hours, immunoassay by thin-layer chromatography from urine samples taken at interview. Sensitivity and specificity were estimated as 0.775 and 0.921, respectively, among the hospital patients.

This validation study is not ideal; recent exposure could genuinely disagree with longer-term use without being a false-positive or a false-negative, and no case-specific estimates were given. However, in the validation study, investigators also noted that sensitivity was higher among hospital patients than among healthy control participants drawn from other visitors to the hospital (0.775 vs 0.688), suggesting that the similar settings of Aliramaji et al. (2015) and the validation study are a strength (although an alternative explanation is that the conditions requiring hospitalization may have increased recent opioid use among regular users). Nonetheless, the sensitivity and specificity in this study are within the range of previous studies of illicit substance use (Harrison et al., 2007).

The values of sensitivity and specificity were used to calculate a bias factor for exposure misclassification, which was assumed to be non-differential, because of the lack of information about bladder cancer status in the validation study. Notably, the bias factor used here is valid when the target parameter is an odds ratio or in situations in which the target parameter estimates an odds ratio (e.g. the risk ratio estimates the odds ratio with a rare outcome), representing a limitation of the bias-factor approach to exposure misclassification. Adjustment for outcome misclassification is not subject to a similar caveat. A further shortcoming of this approach is that some misclassification parameter values will be incompatible with the data (e.g. may result in implausible values for exposure prevalence). The misclassification bias factor was 1.0, which is a result of the observation that non-differential misclassification of a binary exposure will lead to bias away from the null, so that non-differential misclassification will not result in a reduction of the bound of plausible parameter values that are consistent with the data and bias parameters. The selection-bias- and exposure-misclassification-adjusted lower bound relative risk is equal to

$$RR_{app}/(BF_S \times BF_M) = 1.24/(1.40 \times 1.0) = 0.89 \quad \text{(A3.7)}$$

A3.2.5 Exposure misclassification bias adjustment using pseudo-data

Again using the validation data from Rashidian et al. (2017), sensitivity and specificity were estimated as 0.775 and 0.921, respectively. Fox et al. (2021) give a formula for creating pseudo-data from a 2 × 2 table (binary exposure, binary outcome), which is used in the R code

provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat) (Fox et al., 2021). This yielded pseudo-data adjusted for selection bias and exposure misclassification (Table A3.5), which notably resulted in an adjusted odds ratio of 28.2 (which, it should be noted, is verging on implausible and relies on a corrected count of exposed control participants of only 6.2).

A3.2.6 Unmeasured confounding bias adjustment using bias factors

After correcting for selection bias by matching on sex, there will be residual confounding by sex in the study. One approach to this residual confounding is to treat sex as an unmeasured confounder and conduct a bias analysis. Bias analysis for unmeasured confounding through bias factors is operationally similar to that for selection bias, in that parameters for the conditional probability of the unmeasured confounder and the outcome must both be specified. That is, for an unmeasured confounder U_c with discrete levels, prevalence parameters given by $\Pr(U_c = u \mid A^* = a, C = c)$ are required; these are used to quantify the maximal relation between the confounder and exposure and risk parameters given by $\Pr(Y = 1 \mid A^* = a, C = c, U_c = u)$, which are used to quantify the maximal relation between the confounder and the outcome. This approach is quite general, because U_c can be binary, categorical, or continuous; it is identical for many scenarios, and a full distribution of the confounder does not have to be specified.

As with selection bias, the prevalence parameters are used to quantify the maximum risk ratio that contrasts prevalence values across exposure values. For example, if U_c is sex, the prevalence ratio contrasts the prevalence of being male (or female) across levels of exposure, and takes the maximum of those two prevalence ratios. Here, the only parameter of crucial interest (for sex as a binary confounder) is

$$\Pr(U_c = \text{male} \mid A^* = 1, C = c) / \Pr(U_c = \text{male} \mid A^* = 0, C = c) \quad \text{(A3.8)}$$

for which it is assumed that the measured covariates C are not crucial to the problem (e.g. the confounder–exposure relation does not change substantially after adjusting for C, and the necessary parameters can be simplified to $\Pr(U_c = u \mid A^* = a)$.)

These parameters can be drawn from survey data. As in the adjustment for selection bias, parameters are taken from the study by Moradinazar et al. (2020), which is used to present one conceptual issue: for selection bias, exposure from that study is treated as a mismeasured exposure, whereas for confounding bias it is necessarily treated as a gold standard exposure. Regardless, it is unlikely that survey data could be identified using a better measure of opium use than self-report, and an assumption that there is no unmeasured confounding can be much stronger than the assumptions inherent in bias analysis.

To calculate the bias parameters, the first calculation is

$$\Pr(U_c = u \mid A = a) = \Pr(A = a \mid U_c = u) \Pr(U_c = u) / \Pr(A = a) \quad \text{(A3.9)}$$

As before, exposure prevalence by sex can be taken directly from the survey data estimates of $\Pr(A^* = a \mid U_S = \text{male}) = 24.1\%$ and $\Pr(A^* = a \mid U_S = \text{female}) = 2.2\%$. A 1:1 sex ratio yields $\Pr(U_S = u) = 50\%$, and (for example) the marginal probability of misclassified exposure is given by the sex-standardized probability of exposure:

$$\begin{aligned} &\textstyle\sum_u \Pr(U_S = u) \Pr(A^* = 1 \mid U_S = \text{male}) \\ &= 0.5 \times 0.022 + 0.5 \times 0.241 \\ &= 0.1315 \end{aligned} \quad \text{(A3.10)}$$

Table A3.5. Selection-bias- and exposure-misclassification-adjusted pseudo-data, exposure misclassification parameters from Rashidian et al. (2017), and the calculated selection-bias-adjusted risk ratio from Aliramaji et al. (2015)

		Bladder cancer		Case odds/OR
		Case participants	Control participants	
Opium use	Yes	63.49	6.24	10.18
	No	111.51	308.61	0.36
Crude/apparent odds ratio				28.18
Misclassification	Sensitivity	0.775	0.775	
	Specificity	0.921	0.921	

OR, odds ratio.
Sources: Aliramaji et al. (2015); Rashidian et al. (2017).

The full calculation yields a prevalence of being male of 0.92 among the exposed participants and 0.44 among the unexposed participants (and can be used to calculate the same prevalences of being female).

As before, for a binary confounder, the key aspect for the risk parameters is the risk ratio comparing the risk of the outcome for men versus that for women, which was given before as 0.33. These bias parameters yielded an unmeasured-confounding bias factor of 1.53. Thus, the selection-bias-, exposure-misclassification-, and unmeasured-confounding-bias-adjusted lower bound relative risk is equal to

$$RR_{app}/(BF_S \times BF_M \times BF_C)$$
$$= 1.25/(1.40 \times 1.0 \times 1.54) = 0.58 \quad \text{(A3.11)}$$

Thus, a true odds ratio of 0.58 is a lower bound of the true odds ratio that is consistent with the smoking-adjusted odds ratio of 1.25 presented in the study of Aliramaji et al. (2015). That is, after adjustment for selection bias, exposure misclassification, and unmeasured confounding, the study results are consistent with odds ratios as low as 0.58.

A3.2.7 Unmeasured confounding bias adjustment using pseudo-data

The bias parameters used for the pseudo-data approach also include the risk ratio, comparing the risk of outcomes for men versus women and the prevalence of being male (or female), given exposure. These parameters resulted in a selection-bias-, exposure-misclassification-, and unmeasured-confounding-bias-adjusted relative risk of 18.6 (Table A3.6). Notably, this approach assumes that the odds ratio is the same across levels of the covariate (no effect measure modification for the odds ratio), as demonstrated by the sex-specific odds ratios of 18.6.

Unlike the bias-factor approach, the pseudo-data approach uses an apparent relative risk that is adjusted for confounding by smoking. Comparing the crude relative risk with the smoking-adjusted relative risk yields a measured-confounding bias of 2.71/1.25 = 2.17, indicating that the crude estimate is too high. The selection-bias-, exposure-misclassification-, and unmeasured-confounding-bias-adjusted odds ratio is further divided by this bias factor. This yields a final adjusted odds ratio of 8.6, which is adjusted for selection bias, exposure misclassification, unmeasured-confounding bias, and measured-confounding bias. This last calculation ignores the fact that estimates of confounding bias will change on adjustment for selection bias and exposure misclassification bias, but it is relatively simple to implement, and estimates are used directly from the data. Unlike the bias-factor approach, which yielded a worst-case odds ratio estimate of 0.58, the pseudo-data approach provides a best-guess odds ratio estimate of 8.6. These results are consistent with each other because they are interpreted differently. The bias-factor estimate indicates that it is possible (but not necessarily likely) that the positive study result could have occurred due to bias alone. The pseudo-data estimate indicates that the positive study result is nonetheless most likely an underestimate of the true odds ratio.

Table A3.6. Selection-bias-, exposure-misclassification-, and unmeasured-confounding-bias-adjusted pseudo-data and risk ratio derived from Aliramaji et al. (2015) and validation studies noted in the text

		Bladder cancer			
		Women		Men	
		Case participants	Control participants	Case participants	Control participants
Opium use	Yes	1.87	0.52	61.62	5.71
	No	33.42	173.76	78.09	134.85
	Total	35.29	174.28	139.71	140.56
Stratified odds ratio (women)		18.63			
Stratified odds ratio (men)				18.63	
Summary odds ratio					18.63

Source: Aliramaji et al. (2015).

A3.3 Sensitivity analysis

As explained elsewhere in this book (e.g. the multidimensional analysis in Section 4.3.2), it is useful to assess how reasonable departures from chosen parameters may influence results. The following scenarios were assessed using the pseudo-data approach: (i) measured-confounding bias only (to assess the accuracy of correcting for measured confounding as a last step); (ii) no false-positive exposures (often assumed where exposure may carry stigma); (iii) false-positive exposures among case participants only; (iv) false-positive exposures among control participants only; (v) a stronger unmeasured-confounder–outcome relation; (vi) selection bias arising from the recruitment of hospital-based control participants; and (vii) alternative exposure misclassification parameters obtained from a study by Abnet et al. (2004). Notably, point estimates from this sensitivity analysis ranged from 1.28 (false-positive exposures among case participants only) to 15.4 (additional selection bias arising from the recruitment of hospital-based control participants), but none of the point estimates was below the null (Table A3.7).

A3.4 A potential probabilistic multiple-bias analysis strategy

Either the bias-factor approach or the pseudo-data approach could be amenable to a probabilistic bias analysis, wherein the fixed values of the bias parameters given are replaced with values drawn from appropriate distributions (Table A3.2). Examples of such an approach are given elsewhere in this book (e.g. Example 4.21), and there are no additional complications to applying those approaches to multiple-bias analysis, so an explicit example of probabilistic multiple-bias analysis is omitted here. However, the R code provided in Annex 2 (online only; available from: https://publications.iarc.who.int/634#supmat) gives an example of how such an analysis could be carried out using the same functions used to conduct the multiple-bias analysis with pseudo-data discussed in Section A3.2. Crucially, the parameter distributions used in the code were arbitrarily chosen because reasonable parameter distributions could not be obtained for the example in this annex. Nonetheless, the code may be used to facilitate probabilistic bias analysis when reasonable and informative distributions can be specified over the bias parameters.

Table A3.7. Sensitivity analysis results with the pseudo-data approach using alternative bias parameters

	No bias adjustment	Bias adjustment scenario						
		Base analysis	No false-positive exposures	False-positives among cases participants only	False-positives among control participants only	Stronger confounder–outcome relation	Additional selection from hospital-based control participants	Misclassification parameters from Abnet et al. (2004)
Selection bias								
Selection probability, exposed case participants	1	1	1	1	1	1	1	1
Selection probability, unexposed case participants	1	1	1	1	1	1	1	1
Selection probability, exposed control participants	1	0.93	0.93	0.93	0.93	0.93	1.00	1.00
Selection probability, unexposed control participants	1	0.52	0.52	0.52	0.52	0.52	0.52	0.52
Exposure misclassification								
Case sensitivity	1	0.78	0.78	0.78	0.78	0.78	0.78	0.90
Case specificity	1	0.92	1.00	0.92	1.00	0.92	0.92	0.93
Control sensitivity	1	0.78	0.78	0.78	0.78	0.78	0.78	0.90
Control specificity	1	0.92	1.00	1.00	0.92	0.92	0.92	0.93
Unmeasured binary confounder								
RR($U \rightarrow Y$)	1	0.33	0.33	0.33	0.33	0.16	0.33	0.33
Pr($U = 1 \mid$ exposed)	1	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Pr($U = 1 \mid$ unexposed)	1	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Measured-confounding bias	2.17	2.17	2.17	2.17	2.17	2.17	2.17	2.17
Summary odds ratio	1.25	8.58	1.68	1.28	11.24	7.42	15.40	4.95

Sources: Abnet et al. (2004); Aliramaji et al. (2015).

References

Abnet CC, Saadatian-Elahi M, Pourshams A, Boffetta P, Feizzadeh A, Brennan P, et al. (2004). Reliability and validity of opiate use self-report in a population at high risk for esophageal cancer in Golestan, Iran. Cancer Epidemiol Biomarkers Prev. 13(6):1068–70. doi:10.1158/1055-9965.1068.13.6 PMID:15184266

Aliramaji A, Kaseean A, Yousefnia Pasha YR, Shafi H, Kamali S, Safari M, et al. (2015). Age distribution types of bladder cancers and their relationship with opium consumption and smoking. Caspian J Intern Med. 6(2):82–6. PMID:26221505

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

Hadji M, Rashidian H, Marzban M, Naghibzadeh-Tahami A, Gholipour M, Mohebbi E, et al. (2022). Opium use and risk of bladder cancer: a multi-centre case-referent study in Iran. Int J Epidemiol. 51(3):830–8. doi:10.1093/ije/dyac031 PMID:35244716

Harrison LD, Martin SS, Enev T, Harrington D (2007). Comparing drug testing and self-report of drug use among youths and young adults in the general population. Rockville (MD), USA: Office of Applied Studies, Substance Abuse and Mental Health Services Administration, United States Department of Health and Human Services. doi:10.1037/e637122007-001

Mansournia MA, Jewell NP, Greenland S (2018). Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol. 33(1):5–14. doi:10.1007/s10654-017-0325-0 PMID:29101596

Moradinazar M, Najafi F, Jalilian F, Pasdar Y, Hamzeh B, Shakiba E, et al. (2020). Prevalence of drug use, alcohol consumption, cigarette smoking and measure of socioeconomic-related inequalities of drug use among Iranian people: findings from a national survey. Subst Abuse Treat Prev Policy. 15(1):39. doi:10.1186/s13011-020-00279-1 PMID:32503660

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

Smith LH, Mathur MB, VanderWeele TJ (2021). Multiple-bias sensitivity analysis using bounds. Epidemiology. 32(5):625–34. doi:10.1097/EDE.0000000000001380 PMID:34224471

Index of examples

Chapter 1. The role of epidemiology in cancer hazard identification by the *IARC Monographs* programme


Example 1.1. Evaluation of radiofrequency electromagnetic field (RF-EMF) radiation by the *IARC Monographs* 13


Example 1.2. Evaluation of red meat consumption by the *IARC Monographs* 14


Example 1.3. Evaluation of night shift work by the *IARC Monographs* 14


Example 1.4. Evaluation of opium consumption by the *IARC Monographs* 15

Chapter 2. Causal diagrams to evaluate sources of bias


Example 2.1a. Motivation for creating a DAG for red meat consumption 26


Example 2.1b. Motivation for creating a DAG for red meat consumption (continued) 29


Example 2.2. Chains, forks, and colliders used in DAGs 30


Example 2.3a. Forks as depictions of shared causes in DAGs 32


Example 2.3b. Conditioning or blocking of paths in DAGs 32


Example 2.4. Depiction of colliders in DAGs 33

Example 2.5. Example of a collider in a randomized controlled trial among downhill skiers 35


Example 2.6a. A possible DAG for red meat consumption and colorectal cancer 37


Example 2.6b. Backdoor paths in a DAG for red meat consumption and colorectal cancer 38


Example 2.6c. Conditioning to block backdoor paths in a DAG 40


Example 2.7a. Depicting shared causes in a DAG for opium consumption and lung cancer 41


Example 2.7b. Simplifying a DAG for opium consumption and lung cancer to account for confounding .. 42


Example 2.7c. Confounding by backdoor paths in the DAG for opium consumption and lung cancer 43


Example 2.8a. Depiction of non-differential measurement error in a DAG for red meat consumption and colorectal cancer 47


Example 2.8b. Depiction of differential measurement error in a DAG for red meat consumption and colorectal cancer 48


Example 2.9. Depiction of collider bias (by hospitalization) 51


Example 2.10. Selection bias in a DAG of opium consumption and lung cancer 52


Example 2.11. Depiction of collider bias (from loss to follow-up) 53


Example 2.12a. Depiction of signed DAGs 55


Example 2.12b. Using signed DAGs to determine the possible impact of biases 56

Chapter 3. Confounding: a routine concern in the interpretation of epidemiological studies


Example 3.1. Self-matching to control for confounding 69

Example 3.2. An example of a natural experiment to control for confounding: a military conscription lottery 70


Example 3.3. Adjustment for body mass index in studies on red meat consumption and colorectal cancer 71


Example 3.4. Overadjustment as a concern in studies on shift work and cancer 72


Example 3.5. Examining confounding by co-exposures in the workplace 73


Example 3.6. Healthy worker survivor bias 74


Example 3.7. Relative importance of confounders 75


Example 3.8. Negative control outcomes 76

Example 3.9. Indirect adjustment of SMRs to reduce healthy worker biases in aluminium smelting work 77

Example 3.10. Negative control exposures 78

Example 3.11. Using a proxy variable to evaluate confounding in a cohort of Seventh Day Adventist adherents 79


Example 3.12. Restriction to one level of a proxy variable to examine residual confounding 79


Example 3.13. Evidence triangulation to evaluate confounding 80


Example 3.14. Use of bounding to examine confounding scenarios 81


Example 3.15. Bias adjustment to evaluate confounding 82


Example 3.16. The *E*-value to evaluate confounding 83

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

Example 4.1. Linear models for measurement error of protein intake from food frequency questionnaires 92

Example 4.2. Berkson error in an example from blood lead and intelligence quotient testing 92

Example 4.3. Non-differential exposure misclassification when exposure is rare versus when exposure is common 93

Example 4.4. Misclassification from categorizing a continuous exposure variable in workers exposed to crystalline silica 93

Example 4.5. Assessing for varying quality of the interviewee response in assessing tobacco smoking 94

Example 4.6. Assessing for varying quality of interviewer in assessing job histories 94


Example 4.7. Recall bias and knowledge of carcinogenicity 95


Example 4.8. Estimation of the extent of recall bias 95


Example 4.9. Proxy respondents and recall bias in a study of pesticide exposure 96

Example 4.10. Using DAGs to identify recall bias 97

Example 4.11. Negative control exposures to assess recall bias in a study of pesticide exposure 98

Example 4.12. Positive control outcomes to assess exposure misclassification in a study of benzene exposure 98


Example 4.13. Using national statistics to assess recall bias 99


Example 4.14. Recruiting different types of control groups to assess recall bias 99


Example 4.15. Using triangulation to assess recall bias 100


Example 4.16. Analysis of bias from non-differential exposure misclassification 101


Example 4.17. Analysis of bias from differential exposure misclassification 102


Example 4.18. Multidimensional sensitivity analysis 103


Example 4.19. Sensitivity analysis for both confounding and misclassification 104


Example 4.20. Sensitivity analysis for categorical exposure misclassification 105


Example 4.21. Probabilistic bias analysis for exposure misclassification 106


Example 4.22a. Regression calibration for adjustment for measurement error 108


Example 4.22b. Regression calibration for adjustment for measurement error (continued) 108


Example 4.22c. Regression calibration for adjustment for measurement error (continued) 108


Example 4.22d. Regression calibration for adjustment for measurement error (continued) 108


Example 4.22e. Estimating an adjusted confidence interval with regression calibration 108


Example 4.23. Bias adjustment for misclassified categorical exposures 110

Example 4.24. Bayesian model components for non-differential exposure misclassification in a case–control study 113

Example 4.25. Opium use and HNSCC – bias analysis for categorical data 114

Example 4.26. Non-differential outcome misclassification in studies of low-dose ionizing radiation 115

Example 4.27. Non-differential outcome misclassification from underdiagnosis of prostate cancer 115

Example 4.28. Non-differential outcome misclassification of tumour subtypes 116

Example 4.29. Differential outcome misclassification among firefighters 116

Example 4.30. Sensitivity analysis for outcome misclassification 116

Chapter 5. Selection bias and other miscellaneous biases

Example 5.1. Selection bias in a case–control study 126

Example 5.2. Magnitude of selection bias 127

Example 5.3. Assessing bias due to non-response at baseline in an occupational cohort study of flight attendants 128

Example 5.4. Bias due to loss to follow-up in an occupational cohort study of flight attendants 129

Example 5.5. Identifying time zero in an occupational cohort study of flight attendants 130

Example 5.6. Left truncation as a source of selection bias in studies of hormone replacement therapy 130

Example 5.7. Left truncation in a population-based cohort study of breast cancer 131

Example 5.8. Right truncation in the cohort of atomic bomb survivors in Japan 132

Example 5.9. Cross-validation to improve ascertainment of case participants 133
Example 5.10. Potential bias resulting from differential selection of case participants 134
Example 5.11. Potential bias from incomplete case ascertainment of benign tumours..... 134
Example 5.12. Potential selection bias arising from different sources of case ascertainment 135
Example 5.13. Potential bias from excluding people with previous cancer from the study..... 135
Example 5.14. Potential bias arising from inclusion of prevalent cases 136
Example 5.15. Indirect evaluation of potential selection bias from differential participation rates in a case–control study 136
Example 5.16. Evaluating potential Berkson bias in a case–control study..... 137
Example 5.17. Evaluating potential selection bias from recruitment of hospital-based control participants 137
Example 5.18. Recruiting hospital visitors as control participants..... 138
Example 5.19. Triangulation across control groups in a study of titanium dioxide exposure..... 138
Example 5.20a. Potential bias from non-participation in a population-based case–control study 139
Example 5.20b. Demographic variables as surrogates for examining selection bias 140
Example 5.20c. Use of short questionnaires among non-respondents in a case–control study 140

Example 5.20d. Examining the potential for bias from use of proxy interviews 140

Example 5.21. Potential selection bias from differential participation in a case–control study 141

Example 5.22. Potential bias from recruitment of hospital-based control groups..... 141

Example 5.23. Using negative control exposures to examine potential selection bias in a case–control study 142

Example 5.24. Using negative control outcomes to examine potential selection bias in a case–control study 142

Example 5.25. Using dose–response analysis to examine potential selection bias in a case–control study 143

Example 5.26. Using external data on exposure prevalence to examine potential selection bias in a case–control study of pesticide exposure..... 143

Example 5.27. Using external data on exposure prevalence to examine potential selection bias in a case–control study of opium exposure 143

Example 5.28. Using triangulation of findings from different control groups to examine biases in a case–control study of pesticide exposure..... 144

Example 5.29. Comparisons across studies to examine potential biases in case–control studies 144

Example 5.30a. Identifying potential selection bias in a nested case–control study of breast cancer..... 145

Example 5.30b. Quantitative bias analysis to examine potential selection bias in a nested case–control study of breast cancer..... 147

Example 5.30c. Confidence interval estimation when quantifying selection bias for a nested case–control study of breast cancer 148


Example 5.30d. Probabilistic bias analysis to examine potential selection bias in a nested case–control study of breast cancer 149

Example 5.30e. Applying probabilistic bias analysis results to estimated odds ratios in a nested case–control study of breast cancer 149

Example 5.31. Immortal time bias in a registry study related to solar radiation exposure 151

Example 5.32. Examining reverse causation in a case–control study of oesophageal cancer 151

Example 5.33. Examining protopathic bias in case–control studies of opium use and cancer 151

Example 5.34. Evaluating the potential impact of protopathic bias in a study of pancreatic cancer 152

Example 5.35. Inappropriate adjustment for a collider 152

Example 5.36. Modelling using biomarker-based proxies of exposure in a study of herbicide exposure 153

Example 5.37. Differential errors resulting from use of biomarkers in studies of Burkitt lymphoma 154

Chapter 6. Incorporating bias assessments into evidence synthesis

Example 6.1. Selection of key biases for night shift work 163

Example 6.2. Selection of key biases for opium consumption 163

Example 6.3. Specifying key confounders 165

Example 6.4. Assessing exposure misclassification 166

Example 6.5. Identifying selection bias 167


Example 6.6. Bias assessment summary table 168


Example 6.7. Bias assessment summary with few informative studies 169

Example 6.8. Bias assessment summary using triangulation 171


Example 6.9. Multiple-bias analysis 172

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


Example 7.1. Red meat consumption and cancer 180

Example 7.2. Water arsenic concentration and cancer 181

Example 7.3. Use of a negative control outcome in a study of hypertension and cancer 183

Example 7.4. Use of a negative control exposure in a study of maternal alcohol consumption and hypertensive disorders of pregnancy 184

Example 7.5. Indirect methods to evaluate confounding in a cohort of lead smelter workers 185


Example 7.6. Bias from inadequate adjustment for time-varying confounding 186

Example 7.7. Use of g-methods to control for time-varying confounding in a study of titanium dioxide exposure 186


Example 7.8. Explaining assumptions about differential sources of error 188


Example 7.9. Regression calibration to quantify bias due to measurement error 188


Example 7.10. Probabilistic bias analysis to quantify recall bias 189


Example 7.11. Case–case analyses to quantify recall bias 190


Example 7.12. Interphone study on mobile phone use and the risk of brain tumours 192


Example 7.13. Mobile phone use and the risk of uveal melanoma 194


Example 7.14. Selection probabilities in the Interphone study 195


Example 7.15. Loss to follow-up and the association between shift work and breast cancer 197


Example 7.16. Case–control study of opium exposure and oesophageal cancer 199

Index of tools

Chapter 3. Confounding: a routine concern in the interpretation of epidemiological studies

Tool C-1 DAGs and signed DAGs 75

Tool C-2 Negative control outcomes (or exposures) and proxies 76

Tool C-3 Triangulation across studies 78

Tool C-4 Bias adjustment 80

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

Tool E-1 Use of substantive knowledge and DAGs for misclassification 96

Tool E-2 Negative control exposures and positive control outcomes 97

Tool E-3 Examination of exposure information from different sources 98

Tool E-4 Comparisons with external data 98

Tool E-5 Consideration of analysis stratified by index versus proxy interviews 98

Tool E-6 Triangulation using comparisons across studies 99

Tool E-7 Simple bias analysis for exposure misclassification 100

Tool E-8 Multidimensional analysis 100
Tool E-9 Multiple categorical bias analysis 104
Tool E-10 Probabilistic bias analysis 104
Tool E-11 Regression calibration for continuous and categorized measures of exposure 106
Tool E-12 Other methods for quantifying bias 110

Chapter 5. Selection bias and other miscellaneous biases

Tool S-1 Substantive knowledge and DAGs 141
Tool S-2 Negative control exposures 142
Tool S-3 Negative control outcomes 142
Tool S-4 Ad hoc reanalysis of published data 142
Tool S-5 Comparisons with external data 142
Tool S-6 Using several control groups 144
Tool S-7 Assessment of selection bias through comparisons across studies 144
Tool S-8 Selection bias adjustment 145

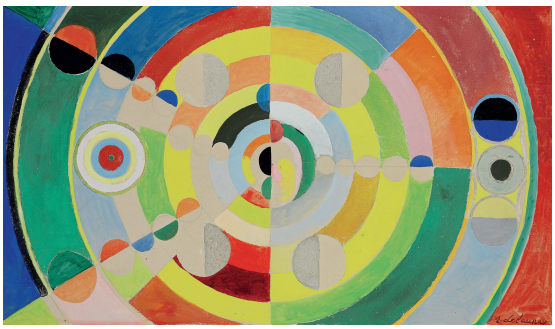
Disclosures of interests

Each participant was asked to declare potentially relevant research, employment, and financial interests that are current or that have occurred during the past 4 years. Minimal interests are not disclosed here, and include stock valued at no more than US$ 1000 overall, grants that provide no more than 5% of the research budget of the expert's organization and that do not support the expert's research or position, and consulting or speaking on matters not before a court or government agency that does not exceed 2% of total professional time or compensation. All other non-publicly funded grants that support the expert's research or position and all consulting or speaking on behalf of an interested party on matters before a court or government agency are disclosed as potentially significant conflicts of interests.

Richard MacLehose reports receiving personal consultancy fees from the Electric Power Research Institute in his capacity as consultant on the impact of confounding by socioeconomic status in the air pollution literature.

David A. Savitz reports receiving personal consultancy fees from Faraci Lange, LLP in his capacity as consultant in cases of exposure to per- and polyfluoroalkyl substances.

Leslie Stayner reports providing expert opinion on behalf of plaintiffs in relation to litigation in cases involving exposure to ethylene oxide and asbestos.



"Relief-disques" by Robert Delaunay

ISBN 978-92-832-2228-6