

Control of Confounding and Reporting of Results in Causal Inference Studies: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals

David J Lederer, MD, MS*¹, Scott C Bell, MBBS, MD, FRACP*², Richard D Branson, MSc, RRT*³, James D Chalmers, MD, PhD*⁴, Rachel Marshall, BSc*⁵, David M Maslove, MD, MS*⁶, David E Ost, MD, MPH*⁷, Naresh M Punjabi, MD, PhD*⁸, Michael Schatz, MD, MS*⁹, Alan R Smyth, MA MBBS MRCP MD FRCPC*¹⁰, Paul W Stewart, PhD*¹¹, Samy Suissa, PhD*¹², Alex A Adjei, MD, PhD¹³, Cezmi A Akdis¹⁴, MD, Élie Azoulay, MD, PhD¹⁵, Jan Bakker, MD, PhD, FCCM, FCCP^{16,17,18}, Zuhair K Ballas, MD¹⁹, Philip G Bardin, FRACP, PhD²⁰, Esther Barreiro, MD, PhD²¹, Rinaldo Bellomo, MD, PhD²², Jonathan A Bernstein, MD²³, Vito Brusasco, MD²⁴, Timothy G Buchman, PhD, MD²⁵, Sudhansu Chokroverty, MD, FRCP²⁶, Nancy A Collop, MD²⁷, James D Crapo, MD²⁸, Dominic A Fitzgerald, MBBS, PhD, FRACP²⁹, Lauren Hale, PhD³⁰, Nicholas Hart, MB, BS, BSc, MRCP, PhD, FFICM³¹, Felix J Herth MD, PhD³², Theodore J Iwashyna, MD, PhD³³, Gisli Jenkins, BM, PhD³⁴, Martin Kolb, MD, PhD³⁵, Guy B. Marks, MB, BS, PhD³⁶, Peter Mazzone, MD, MPH³⁷, J Randall Moorman, MD³⁸, Thomas M Murphy, MD³⁹, Terry L Noah, MD⁴⁰, Paul Reynolds, MBBS, PhD, MD⁴¹, Dieter Riemann, PhD⁴², Richard E Russell, MD, PHD^{43,44}, Aziz Sheikh, MD, MSc,⁴⁵ Giovanni Sotgiu, MD, PhD⁴⁶, Erik R Swenson, MD⁴⁷, Rhonda Szczesniak, PhD⁴⁸, Ronald Szymusiak, PhD⁴⁹, Jean-Louis Teboul, MD, PhD⁵⁰, Jean-Louis Vincent, MD, PhD⁵¹

ORCID IDs: 0000-0001-5258-0228 (DJL), 0000-0001-8651-7139 (SCB), 0000-0002-0912-3360 (RDB), 0000-0002-0765-7158 (DMM), 0000-0001-8020-019X (CAA), 0000-0003-2236-7391 (JB), 0000-0002-8976-8053 (GBM), 0000-0002-5772-1648 (JRM).

¹Departments of Medicine and Epidemiology, Columbia University Irving Medical Center, New York, New York; ²Department of Thoracic Medicine, The Prince Charles Hospital, Brisbane, Australia; ³Department of Surgery, University of Cincinnati, Cincinnati, Ohio; ⁴University of Dundee, Dundee, Scotland; ⁵London, England; ⁶Department of Medicine, Queen's University Kingston, Ontario, Canada; ⁷Department of Pulmonary Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas; ⁸Division of Pulmonary and Critical Care Medicine Johns Hopkins University, Baltimore, Maryland; ⁹Department of Allergy, Kaiser Permanente Medical Center, San Diego, California; ¹⁰Division of Child Health, Obstetrics & Gynecology, ³⁴Department of Experimental Medicine, University of Nottingham, Nottingham, England; ¹¹Department of Biostatistics, ⁴⁰Department of Pediatrics, Division of Pulmonology, University of North Carolina, Chapel Hill, North Carolina; ¹²Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada; ¹³Department of Oncology, Mayo Clinic, Rochester, Minnesota; ¹⁴Swiss Institute of Allergy and Asthma Research, University of Zurich, Davos, Switzerland; ¹⁵St Louis Hospital, University of Paris, Paris, France; ¹⁶Department of Pulmonology and Critical Care, Columbia University Irving Medical Center and NYU Langone Health, New York; ¹⁷Department of Intensive Care Adults, Erasmus MC University Medical Center, Rotterdam, Netherlands; ¹⁸Department of Intensive Care, Pontificia Universidad Católica de Chile, Santiago, Chile; ¹⁹Department of Internal Medicine, University of Iowa and the Iowa City VA Medical Center, Iowa City, Iowa; ²⁰Monash Lung & Sleep, Monash Hospital and University, Melbourne, Victoria, Australia; ²¹Pulmonology Department-Muscle & Lung Cancer Research Group, Research Institute of Hospital del Mar and CIBERES (ISC-III), Barcelona, Spain; ²²Department of Intensive Care Medicine, Austin Hospital and University of

Melbourne, Melbourne, Australia; ²³Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio; ²⁴Department of Internal Medicine, University of Genoa, Genoa, Italy; ²⁵Departments of Surgery, Anesthesiology, and Biomedical Informatics, ²⁷Departments of Medicine and Neurology, Emory University School of Medicine, Atlanta, Georgia; ²⁶JFK NJ Neuroscience Institute, Hackensack Meridian Health-JFK Medical Center, Edison, New Jersey; ²⁸Department of Medicine, National Jewish Hospital, Denver, Colorado; ²⁹The Children's Hospital at Westmead, Sydney Medical School, University of Sydney, Sydney, Australia; ³⁰Department of Family, Population and Family Medicine, Stony Brook University, Stony Brook, New York; ³¹Lane Fox Respiratory Service, Guy's & St Thomas' Hospital, London, United Kingdom; ³²Department of Pneumology and Critical Care Medicine, University of Heidelberg, Heidelberg, Germany; ³³Department of Internal Medicine, University of Michigan, VA Center for Clinical Management Research, VA Ann Arbor Health System, Ann Arbor, Michigan; ³⁵Department of Medicine, McMaster University, Hamilton, Ontario, Canada; ³⁶South Western Sydney Clinical School, University of New South Wales, Sydney, Australia; ³⁷Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, Ohio; ³⁸Departments of Medicine, Physiology, Engineering University of Virginia, Charlottesville, Virginia; ³⁹Division of Pulmonary & Sleep Medicine, MassGeneral Hospital for Children, Harvard Medical School, Boston, Massachusetts; ⁴¹Department of Thoracic Medicine, Royal Adelaide Hospital, University of Adelaide, Adelaide, Australia; ⁴²Department of Psychiatry & Psychotherapy, Freiburg University Medical Center, Freiburg, Germany; ⁴³Senior Clinical Researcher Nuffield Dept of Medicine, University of Oxford; ⁴⁴Consultant Physician, Lymington New Forest Hospital, Hampshire, UK; ⁴⁵Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, Scotland, United Kingdom; ⁴⁶Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Sassari, Italy; ⁴⁷Department of Medicine, University of Washington, VA Puget Sound Health Care System, Seattle, Washington; ⁴⁸Division of Biostatistics & Epidemiology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati, Cincinnati, Ohio; ⁴⁹Departments of Medicine and Neurobiology, David Geffen School of Medicine at UCLA, Los Angeles, California; ⁵⁰CHU Bicêtre, Le Kremlin-Bicêtre, France; ⁵¹Bicetre Hospital, AP-HP, University Paris South, Le Kremlin-Bicetre, France

*Writing group

Corresponding Author:

David J Lederer, MD, MS
 Associate Professor of Medicine and Epidemiology
 Columbia University Irving Medical Center
 161 Fort Washington Ave, Rm 3-321A
 New York, NY 10032; Phone: 212-305-8203
DL427@cumc.columbia.edu

Funding: None

Disclaimer: The views and recommendations made in this document do not represent the official position of any publisher or professional medical society. This document has not been

endorsed by any publisher or other official entity. This document does not necessarily represent the official views of the US Government or Department of Veterans Affairs.

Running Head: Guidance for Authors

MeSH Headings: Epidemiology, Causality, Research Design, Confounding Factors

Word Count: 3,119

The 21st century has brought with it a welcome call for increased rigor in observational research methods (1, 2). It is not that observational research methods are inherently flawed – they are not (3, 4). Observational studies can contribute valuable evidence supporting causal associations when designed and conducted using rigorous methods. The “flaws” are a result of reliance on outdated methodology, inadequate attention to threats to validity (such as confounding), opaque reporting of results, lack of replication, and a failure to interpret findings within the context of the limitations of observational research methodology.

Aware of this situation and influenced by our experience as journal editors, we convened an *ad hoc* group of 47 editors of 35 respiratory, sleep, and critical care journals to offer guidance to authors, peer reviewers, and researchers on the design and reporting of observational causal inference studies. This guidance takes the form of a call for investigators to consider making major changes to their approach to such studies. This document represents our current best understanding of approaches to causal inference, an active area of research. We anticipate that best practice in this, as in any scientific endeavor, will continue to evolve, requiring this document to be updated every 5 to 10 years. We believe these changes will increase the rigor, validity, and value of the work we publish in our journals.

What Is Causal Inference?

We first wish to make a distinction between causal inference and prediction modeling. Causal inference is the examination of causal associations to estimate the causal effect of an exposure on an outcome. We use causal inference to answer questions about etiology: Does long-term

exposure to traffic-related air pollution promote obstructive sleep apnea in non-obese adults? Does caffeine intake protect against pulmonary arterial hypertension? Do anti-depressants reduce the risk of the ARDS in adults with community-acquired pneumonia? Both experimental studies (e.g., randomized clinical trials) and observational studies (e.g., cohort, case-control, and cross-sectional studies) can be used to examine causal associations. We encourage authors to design observational studies that emulate the clinical trial they would have designed to answer the causal question of interest (5, 6). Causal inference studies require a clearly articulated hypothesis, careful attention to minimizing selection and information bias, and a deliberate and rigorous plan to control confounding. The latter is addressed in detail later in this document.

Prediction models are fundamentally different than those used for causal inference (7). Prediction models use individual-level data (predictors) to estimate (predict) the value of an outcome. For example, one might wish to predict an adult's 10-year risk of developing lung cancer. Investigators might use machine learning methods, penalized estimation, or one of many other available methods to develop a prediction model using a dataset containing both the predictors of interest and lung cancer event data. A risk score calculator (or other clinically useful tool) could then be developed, validated, disseminated, and implemented in practice. This document does not address development, validation, or reporting of prediction model.

With this background, we offer three key principles to guide authors in the analysis and reporting of causal inference studies (Table 1):

Key Principle #1: Causal Inference Requires Careful Consideration of Confounding

Herein, we focus on how one should define and select confounders in observational studies that attempt to make causal inferences. Based on our experience, we have identified 5 approaches commonly used by authors (Table 1). Only two of these methods (the “historical” approach and causal modeling), however, aid in causal inference. The others, those based on statistical hypothesis testing or model fit, do not. We detail each approach below:

Historical Approach to Defining a Confounder

A confounder has long been defined as any third variable that is associated with the exposure of interest, is a cause of the outcome of interest, and does not reside in the causal pathway between the exposure and outcome (Figure 1A) (8). We find this definition reasonable, and we regard it as an acceptable approach to address confounding in studies of causal inference. Importantly, as clarified later, we expect authors to *purposefully* select variables that plausibly fit these criteria based on prior knowledge rather than selecting those variables associated with the exposure or outcome using the available data.

Using Causal Models to Identify Confounding

While the historical approach described above is acceptable for simple causal structures, it is often inadequate to describe the more commonly encountered causal networks. Hence, we urge authors to consider using causal models when testing causal associations.

The scientific, mathematical, and theoretical underpinnings of causal inference, developed by Judea Pearl, James Robins, Miguel Hernán, and others, have evolved sufficiently to permit the everyday use of causal models (9-17). Causal models can be represented visually using directed acyclic graphs (DAGs). A DAG is a graph in which unidirectional arrows are used to represent known causal effects (based on prior knowledge). While investigators often feel some discomfort in deciding what causal effects do and do not exist based on prior knowledge, the advantage of this approach is that it makes these assumptions explicit (and hence transparent). In fact, all other methods of controlling for confounding involve implicit assumptions about causal effects, which are not transparent to the reader.

Four simple DAGs are shown in Figure 1. Within a DAG, a “path” is a set of arrows connecting any two variables (regardless of arrow direction). The causal path of interest is the hypothesized association between the exposure and outcome. A “back-door path” is an alternate path between the exposure and the outcome. *Confounding* is defined as the presence of at least one “open” back-door path between exposure and outcome. Variables that naturally open back-door paths are called *confounders*. An association will exist between any two variables connected by an open path. When an investigator “controls” for a confounder, the back-door path will be “closed,” and the association between the exposure and outcome will no longer be observed.

As an example, suppose an investigator is testing whether exercise is associated with a reduced risk of lung cancer. In Figure 1A, there is one causal path: exercise \rightarrow lung cancer, and one back-door path: exercise \leftarrow smoking \rightarrow lung cancer. This open back-door path indicates the presence of confounding, and therefore smoking is a confounder of the causal association

between exercise and lung cancer. Note that we define a confounder here as a variable that, when controlled for, closes a back-door path.

When more than one variable lies along a back-door path, control of a single confounder on the path is sufficient to close the back-door path. In a fully developed DAG with many paths, control of a small number of variables (a “minimum set” of confounders) will often close all back-door paths. We recommend using this approach in causal inference studies. DAGitty.net offers authors a simple interface with which to construct DAGs and identify back-door paths and minimum sets of confounders (18).

Figure 1B adds another type of variable -- a mediator -- to the DAG. A mediator is a variable that lies along the causal path (not a back-door path) between the exposure and disease. Mediators are, of course, of great interest, since they are causes and mechanisms of disease. In Figure 1B, the mediator is “immune function.” At least some of the causal effect of exercise on lung cancer risk is mediated by exercise-induced changes: exercise → immune function → lung cancer. A path that includes a mediator is often called an indirect effect or *indirect causal path*. In contrast, the arrow directly connecting exercise and lung cancer represents the *direct causal effect* of exercise on lung cancer not due to changes in immune function.

Mediators naturally leave the indirect causal path open. Control of a mediator (through adjustment or other means) will close the indirect causal path, preventing or limiting the ability to observe an association between the exposure and outcome (if indeed one exists). Mediators therefore require special attention (if they are to be examined at all) and should not be treated

as confounders. Use of a DAG can aid investigators in identifying mediators, thereby avoiding control of these variables in testing causal effects.

A discussion of “collider bias” further illustrates the value of using DAGs. A “collider” is a variable with two or more antecedent causes that lie within a pathway of interest. A collider can be identified on a DAG when two arrows along a path both point to a variable (Figure 1C). When both the exposure and outcome are causes of the collider, one may be tempted to control for the collider. However, *colliders naturally block back-door paths*. Controlling for a collider will open the back-door path, thereby *introducing* confounding.

For example, in Figure 1C we are interested in testing the causal association between shift work and obstructive sleep apnea. We might be tempted to control for sleepiness, since both shift work and obstructive sleep apnea cause sleepiness. However, sleepiness is a collider that naturally blocks the back-door path of shift work \rightarrow sleepiness \leftarrow obstructive sleep apnea. Controlling for sleepiness would open this back-door path, introducing confounding.

To clarify, imagine that, in reality, shift work is not a cause of obstructive sleep apnea. If we encountered a sleepy person with obstructive sleep apnea, their sleep apnea would likely be the cause of their sleepiness, and therefore they would be less likely to be a shift worker. Conversely, if we encountered a sleepy shift worker, it is likely that shift work is the cause of their sleepiness rather than obstructive sleep apnea. We would therefore observe that sleep apnea occurs less commonly among shift workers, and thus report an inverse association. This confounded association results from conditioning on a collider (in this case, by only examining sleepy people). The same bias would occur if we were to adjust for sleepiness using a regression model.

Collider bias may also be present when neither the exposure nor the outcome is a direct cause of the collider variable. An example is “M-bias,” named after the shape of the DAG (Figure 1D) (19). In this example, we are testing the causal association between chronic beta-blocker use and the risk of developing ARDS. We might be tempted to adjust for the presence of auscultatory crackles at hospital admission, since: (a) heart failure leads to both chronic beta-blocker therapy and crackles; and (b) pneumonia causes both ARDS and crackles. These relationships may lead us to believe that “crackles” is a confounder, whereas in reality it is not. Instead, as Figure 1D shows, “crackles” is a collider on the back-door path of chronic beta-blocker therapy \leftarrow heart failure \rightarrow crackles \leftarrow pneumonia \rightarrow ARDS. Adjusting for the presence of crackles opens this back-door path, introducing confounding. Ignoring the presence of crackles would be the right thing to do.

We encourage investigators that wish to control for variables that do not close a back-door path to ensure that these additional variables are neither mediators nor colliders.

DAGs do come with limitations. They are non-parametric by nature. The directionalities of effects are not always known. DAGs are prone to misspecification when there is a lack of strong background information, and constructing a DAG can be challenging, with even small errors potentially leading to incorrect inferences. Despite these limitations, DAGs lay bare the assumptions made by the investigators, which can then be identified and corrected more readily during pre- and post-publication peer review than through more opaque methods.

This brief document cannot provide a detailed discussion of causal inference, but we hope that these examples encourage authors to consider using causal models in their research. We refer authors to a number of excellent resources on the topic (Table 2).

Variable Selection Methods That Do Not Adequately Control for Confounding

P-value-based and model-based variable selection methods (including forward, backward, and stepwise selection) should not be used for causal inference. These approaches ignore the causal structure underlying the hypothesis and therefore do not adequately control for confounding. Confounders and colliders are treated similarly. Methods relying on model fit or related constructs (such as r^2 , Akaike information criterion, and Bayesian information criterion) also have no relevance to causal inference. These methods rely heavily on the available data, in which causal relationships may or may not have been captured and may or may not be evident. Specification of the model and the arbitrary variables included in any particular model will drive observed associations with the outcome.

Selection of variables that, when included in a model, change the magnitude of the effect estimate of the exposure of interest should not be used to identify confounders, for the reasons discussed above.

Identification of multiple “independent predictors” (“winners”) through purposeful or automated variable selection is an unacceptable approach for testing causal associations. If the authors have hypotheses about each variable, then a separate model for each variable should be generated using one of the above preferred approaches. Alternatively, a prediction model could be developed, if prediction, rather than causal inference, is the goal of the analysis.

“Table 2 Fallacy”

Causal models are typically designed to test an association between a single exposure and an outcome. The additional independent variables in a model (often called “covariates”) serve to

control for confounding. The observed associations between these covariates and the outcome have not been subject to the same approach to control of confounding as the exposure. Therefore, residual confounding and other biases often heavily influence these associations. This situation is known as “Table 2 Fallacy,” a term arising from the practice of presenting effect estimates for all independent variables in “Table 2” (20). We strongly caution authors to avoid presenting these effect estimates in the primary manuscript.

Causal Association, Causal Effect, and Claiming Causality

Readers may find it unusual that we are using the word “causal” to describe observed associations. When examining associations in observational causal inference studies, the intention is always to seek evidence to support (or refute) a true causal effect of the exposure on the outcome. Of course, we often cannot establish these causal effects from any single study. Yet, by acknowledging the intent, it is reasonable to use the label “causal association” (but not “causal effect”) to describe findings arising from an observational study.

We therefore caution authors that claims of causality should be avoided without substantial evidence of a true causal effect, as espoused by Bradford Hill and further developed by John Ioannidis (21, 22). It is reasonable to use the term “effect estimate” when referring to a causal association in an observational study, but assertions that an exposure has an “effect” or “impact” on the outcome, or that the exposure “protects against” or “promotes” the outcome should not be made.

A Note on Methods to Control for Confounding

Investigators may control for confounding either in the design or analysis of a study.

Randomization to exposure, use of an instrumental variable, weighted regression via propensity-scores, adjustment using multivariable regression, stratification on a confounder, conditioning enrollment on a confounder (restriction), and matching on a confounder are common methods (4). We do not make recommendations for or against any of these methods.

Key Principle #2: Interpretation of Results Should Not Rely on the Magnitude of P-Values

In recent years, the merits of the p-value in causal inference have been questioned (23-26). P-values are frequently misinterpreted and misused (27). Although some disagree (28), they provide no information about the magnitude, direction, or clinical importance of an association. Accordingly, we recommend that P-values only rarely be presented in isolation (exceptions may include “omics” studies and tests for interaction). Effect estimates and measures of precision (e.g., confidence intervals or credible intervals) should be presented in addition to (or in place of) p-values.

We recommend interpreting the variability around an effect estimate when making conclusions about causal associations. For example, a rate ratio of 2.1 with a confidence interval of 0.97 to 4.2 and a corresponding p-value of 0.10 should not be reported as “no association,” since a rate ratio as large as 4.2 has not been plausibly excluded, and, at least within the study sample, an association was indeed observed. Instead, a statement such as

“The exposure was associated with a 2.1-fold increased rate of the outcome (95% confidence interval 0.97 to 4.2), but this estimate is imprecise” would be sufficient. In this example, the point- and interval-estimates are informative, yet (not surprisingly) the hypothesis test was inconclusive. Similarly, we recommend against using the vague labels “significant” and “non-significant,” which lead readers (and authors) to implicitly conclude that an association is present or absent. Use of the unqualified word “significant” tends to blur the important distinction between statistical significance and clinical significance. We favor simply reporting the quantitative findings as indicated above. The clinical, mechanistic, or biological interpretations of effect sizes provide greater value and should be used in place of these labels.

Key Principle #3: Results Should Be Presented in a Granular and Transparent Fashion

The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement, published in 2007, provides clear and valuable guidance on the reporting of results of human observational studies that test causal associations (29). We strongly recommend that authors adhere to the STROBE statement when reporting results, including the detailed guidance provided in the STROBE “Explanation and Elaboration” Document (30). In particular, when applicable, results should be presented in Tables modeled after those in Sections 15 and 16 of the STROBE “Explanation and Elaboration” Document (30), with the following in mind:

- In cohort studies, tabular presentation of results should include the number of events, person-time, incidence rates, and unadjusted and adjusted incidence rate ratios for each exposure level.
- In cross-sectional studies, tabular presentation of results should include the number of events, prevalences, and unadjusted and adjusted prevalence ratios for each exposure level.
- In case-control studies, tabular presentation of results should include the number and percent exposed for cases and controls separately, and unadjusted and adjusted odds ratios for each case group.

We encourage authors to take a thoughtful and careful approach to the visual presentation of quantitative results (31). When possible, presentation of individual data points should accompany measures of central tendency and variation. The “data-ink ratio” should be maximized by avoiding unnecessary lines, grids, and text (31). Abbreviations should be used sparingly. Continuous data should not be presented in bar charts with standard error bars (“plunger plots”) (32, 33). Authors should use color-blind friendly palettes.

Final Comment to Our Authors

This document is intended to provide firm guidance rather than absolute rules; to raise the rigor of the work reported in our journals; to improve the communication of research findings; to enhance the value and validity of the science in our field; to aid in replication; and, most

importantly, to improve the health of those living with respiratory disease, sleep disorders, and critical illness.

References

1. Prasad VK, Cifu AS. Ending medical reversal : improving outcomes, saving lives. Baltimore: Johns Hopkins University Press; 2015.
2. Ioannidis JP. Why Most Clinical Research Is Not Useful. *PLoS Med* 2016; 13: e1002049.
3. Kitsios GD, Dahabreh IJ, Callahan S, Paulus JK, Campagna AC, Dargin JM. Can We Trust Observational Studies Using Propensity Scores in the Critical Care Literature? A Systematic Comparison With Randomized Clinical Trials. *Crit Care Med* 2015; 43: 1870-1879.
4. Gershon AS, Jafarzadeh SR, Wilson KC, Walkey AJ. Clinical Knowledge from Observational Studies: Everything You Wanted to Know but Were Afraid to Ask. *Am J Respir Crit Care Med* 2018 (in press).
5. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016; 183: 758-764.
6. Hernan MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. *Epidemiology* 2011; 22: 290-291.
7. Shmueli G. To Explain or to Predict? *Stat Sci* 2010; 25: 289-310.
8. Rothman KJ, Lash TL, Greenland S. Modern Epidemiology, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008. p. 195.
9. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003; 14: 300-306.
10. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999; 10: 37-48.
11. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004; 15: 615-625.
12. Morabia A. History of the modern epidemiological concept of confounding. *J Epidemiol Community Health* 2011; 65: 297-300.
13. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009; 20: 488-495.
14. Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. *Respirology* 2014; 19: 303-311.
15. Wunsch H, Linde-Zwirble WT, Angus DC. Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care* 2006; 21: 1-7.
16. Pearl J. Causality: models, reasoning, and inference. New York, NY: Cambridge University Press; 2009.
17. Pearl J, Mackenzie D. The Book of Why. New York, NY: Basic Books; 2018.

18. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol* 2016; 45: 1887-1894.
19. Ding P, Miratrix LW. To Adjust or Not to Adjust? Sensitivity Analysis of M-Bias and Butterfly-Bias. *J Causal Infer* 2015; 3: 41-57.
20. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* 2013; 177: 292-298.
21. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965; 58: 295-300.
22. Ioannidis JP. Exposure-wide epidemiology: revisiting Bradford Hill. *Stat Med* 2016; 35: 1749-1762.
23. Benjamin D, Berger J, Johannesson M, Nosek B, Wagenmakers EJ, Berk R, Bollen K, Brembs B, Brown L, Camerer C, Cesarini D, Chambers C, Clyde M, Cook T, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field A, Forster M, George E, Gonzalez R, Goodman S, Green E, Green D, Greenwald A, Hadfield J, Hedges L, Held L, Hua Ho T, Hoijsink H, Hruschka D, Imai K, Imbens G, Ioannidis J, Jeon M, Jones J, Kirchner M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell S, McCarthy M, Moore D, Morgan S, Munafó M, Nakagawa S, Nyhan B, Parker T, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt F, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts D, Winship C, Wolpert R, Xie Y, Young C, Zinman J, Johnson V. Redefine statistical significance. *Nature Human Behaviour* 2017; 2: 6-10.
24. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA* 2018; 319: 1429-1430.
25. Lakens D, Adolfs FG, Albers CJ, Anvari F, Apps MAJ, Argamon SE, Baguley T, Becker RB, Benning SD, Bradford DE, Buchanan EM, Caldwell AR, Van Calster B, Carlsson R, Chen S-C, Chung B, Colling LJ, Collins GS, Crook Z, Cross ES, Daniels S, Danielsson H, DeBruine L, Dunleavy DJ, Earp BD, Feist MI, Ferrell JD, Field JG, Fox NW, Friesen A, Gomes C, Gonzalez-Marquez M, Grange JA, Grieve AP, Guggenberger R, Grist J, van Harmelen A-L, Hasselman F, Hochard KD, Hoffarth MR, Holmes NP, Ingre M, Isager PM, Isotalus HK, Johansson C, Juszczak K, Kenny DA, Khalil AA, Konat B, Lao J, Larsen EG, Lodder GMA, Lukavský J, Madan CR, Manheim D, Martin SR, Martin AE, Mayo DG, McCarthy RJ, McConway K, McFarland C, Nio AQX, Nilsson G, de Oliveira CL, de Xivry J-JO, Parsons S, Pfuhl G, Quinn KA, Sakon JJ, Saribay SA, Schneider IK, Selvaraju M, Sjoerds Z, Smith SG, Smits T, Spies JR, Sreekumar V, Steltenpohl CN, Stenhouse N, Świątkowski W, Vadillo MA, Van Assen MALM, Williams MN, Williams SE, Williams DR, Yarkoni T, Ziano I, Zwaan RA. Justify your alpha. *Nature Human Behaviour* 2018; 2: 168-171.
26. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *Am Stat* 2016; 70: 129-133.
27. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016; 31: 337-350.

28. Kulldorff M, Graubard B, Velie E. The P-value and P-value function. *Epidemiology* 1999; 10: 345-347.
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573-577.
30. Vandenbroucke JP, von Elm E, Altman DG, Gotsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147: W163-194.
31. Tufte ER. The visual display of quantitative information. Cheshire, Conn.: Graphics Press; 2001.
32. Drummond GB, Vowler SL. Show the data, don't conceal them. *Br J Pharmacol* 2011; 163: 208-210.
33. Rockman HA. Great expectations. *J Clin Invest* 2012; 122: 1133.

Figure Legends

Figure 1: Directed acyclic graphs illustrating (A) confounding, (B) mediation, (C) collider bias, and (D) M-bias. Each arrow represents a causal effect. In (A), the blue arrows represent an open back-door path: exercise \leftarrow smoking \rightarrow lung cancer. “Smoking” is a confounder that naturally leaves the back-door path open. Controlling for “smoking” will close the back-door path, eliminating confounding through this path. In (B), the black arrows all represent the direct and indirect causal paths of interest. “Immune function” partially mediates the association between exercise and lung cancer: exercise \rightarrow immune function \rightarrow lung cancer. Control of “immune function” would be inappropriate, since it would partially close the causal path, attenuating the observed association between exercise and lung cancer. In (C), the orange arrows represent a closed back-door path: shift work \rightarrow sleepiness \leftarrow obstructive sleep apnea. “Sleepiness” is a collider that naturally leaves the back-door path closed. Control of “sleepiness” would open the back-door path, introducing confounding through this path. In (D), the orange arrows represent a closed back-door path: chronic beta-blocker therapy \leftarrow heart failure \rightarrow crackles \leftarrow pneumonia \rightarrow acute respiratory distress syndrome. “Crackles” is a collider that naturally leaves the back-door path closed. Control of “crackles” would open the back-door path, introducing confounding through this path.

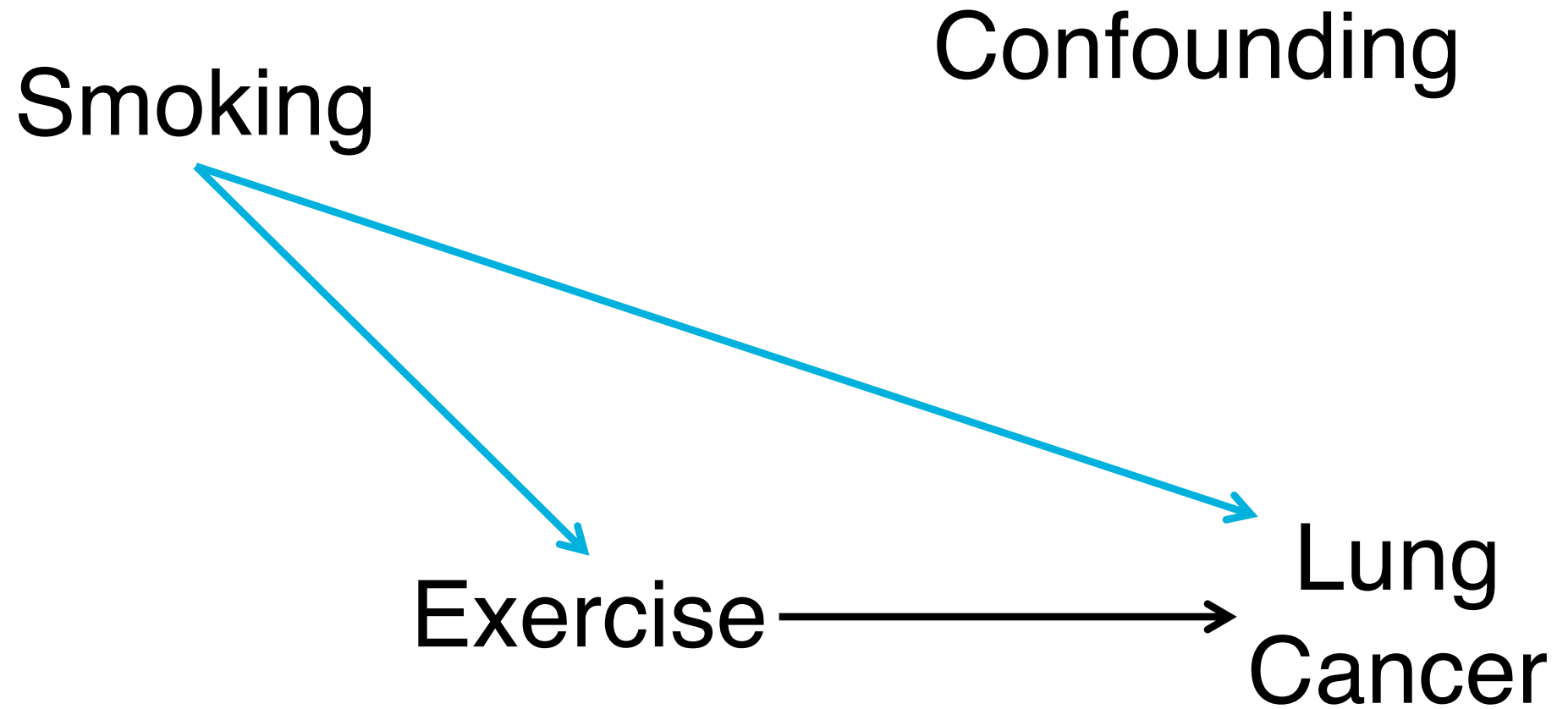
Table 1: Key Principles

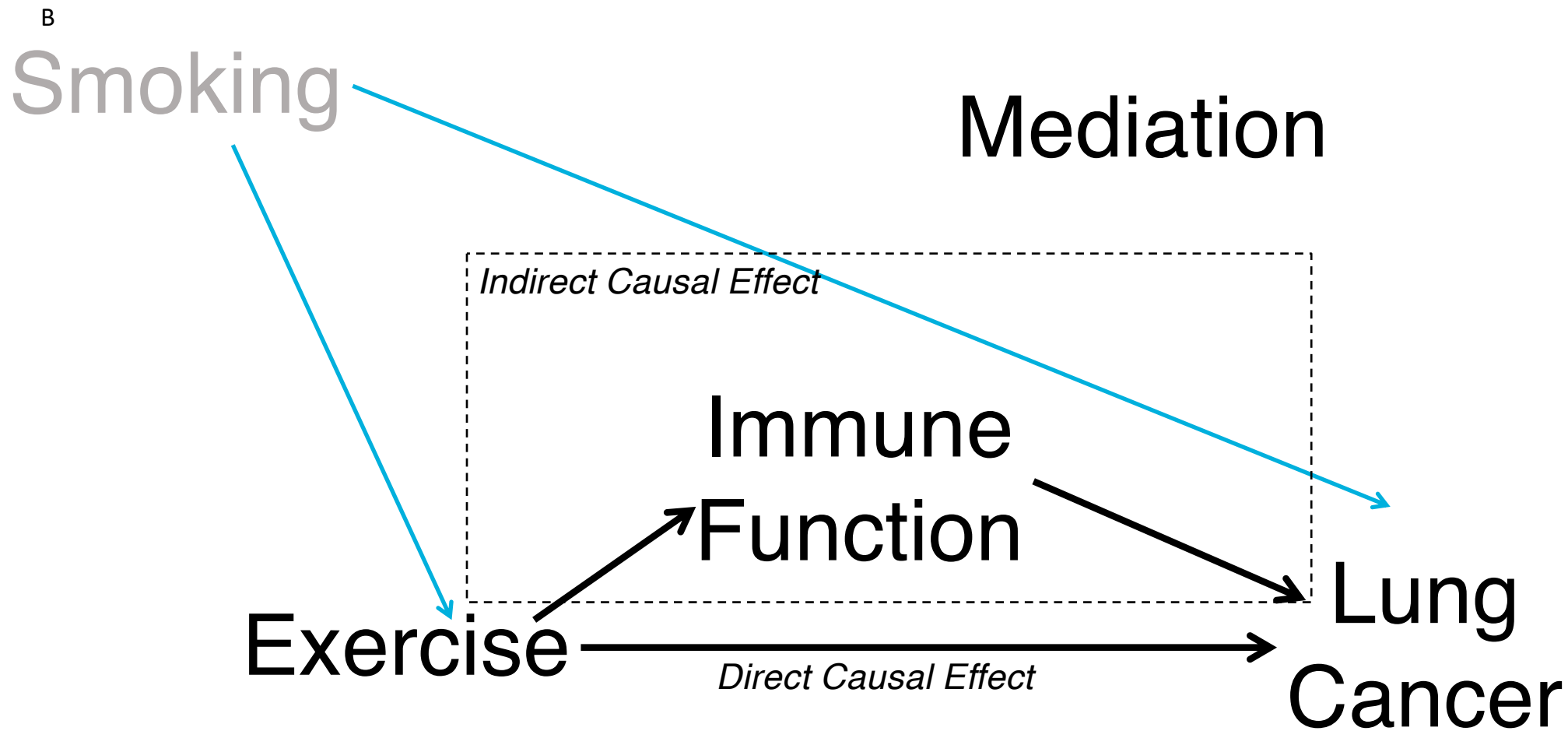
| |
|--|
| <p>Key Principle #1: Causal inference requires careful consideration of confounding</p> <ul style="list-style-type: none"> • Preferred variable selection methods <ol style="list-style-type: none"> 1. Historical confounder definition with purposeful variable selection 2. Causal models using directed acyclic graphs • Variable selection methods that do not adequately control for confounding <ol style="list-style-type: none"> 3. P-value or model-based methods 4. Methods based on beta-coefficient changes 5. Selection of variables to identify “independent predictors” • Do not present all of the effect estimates from a model designed to test a single causal association (“Table 2 fallacy”). |
| <p>Key Principle #2: Interpretation of results should not rely on the magnitude of p-values</p> <ul style="list-style-type: none"> • P-values should rarely be presented in isolation. • Present effect estimates and measures of variability with or without p-values. • Variability around effect estimates should inform conclusions. • A conclusion of “no association” should require exclusion of meaningful effect sizes. • Avoid the word “significant” in favor of more specific language. |
| <p>Key Principle #3: Results should be presented in a granular and transparent fashion.</p> <ul style="list-style-type: none"> • Use the STROBE statement and checklist. • Model tables after the STROBE “Explanation and Elaboration” Document (Reference #31). • Visual presentation of quantitative results <ul style="list-style-type: none"> ○ Present individual data points when possible. ○ Avoid excessive lines, text, grids, and abbreviations. ○ Continuous data should not be presented in bar charts with standard error bars (“plunger plots”). ○ Use color-blind friendly palettes |

Table 2: Causal Inference Resources

| |
|--|
| <p>Books</p> <ul style="list-style-type: none"> • Pearl, J. (2018). The Book of Why: The New Science of Cause and Effect. Basic Books; 1st edition. • Pearl J. (2009). Causality: Models, Reasoning and Inference. Cambridge University Press; 2nd edition. • Hernán MA, Robins JM (2018). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. (Available at: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/) |
| <p>Articles</p> <ul style="list-style-type: none"> • Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. <i>Epidemiology</i> 1999;10(1):37-48. • Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. <i>Epidemiology</i>. 2003;14(3):300-6. • Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. <i>Epidemiology</i> 2004;15(5):615-25. • Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. <i>Epidemiology</i>. 2009;20(4):488-95. • Morabia A. History of the modern epidemiological concept of confounding. <i>J Epidemiol Community Health</i>. 2011;65(4):297-300. • Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. <i>Respirology</i> 2014;19(3):303-11. • Hernán MA. The C-Word: Scientific euphemisms do not improve causal inference from observational data. <i>Am J Public Health</i>. 2018;108(5):616-619. |
| <p>Websites</p> <ul style="list-style-type: none"> • An online course about causal inference and directed acyclic graphs: https://www.edx.org/course/causal-diagrams-draw-assumptions-harvardx-ph559x. Causal Diagrams: Draw Your Assumptions Before Your Conclusions. By Miguel Hernán. • A web-based environment for creating directed acyclic graphs: http://dagitty.net. Textor J, van der Zander B, Gilthorpe MK, Liskiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. <i>International Journal of Epidemiology</i> 45(6):1887-1894, 2016. |

A





c

**Collider
Bias**

Sleepiness

Shift work

**Obstructive
Sleep Apnea**



D

Heart Failure

Pneumonia

M-Bias

Crackles

Beta-Blocker
Use

ARDS

