# PERSPECTIVE

# 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[1,2]*, Scott C. Bell[3]*, Richard D. Branson[4]*, James D. Chalmers[5]*, Rachel Marshall[6]*, David M. Maslove[7]*, David E. Ost[8]*, Naresh M. Punjabi[9]*, Michael Schatz[10]*, Alan R. Smyth[11]*, Paul W. Stewart[12]*, Samy Suissa[13]*, Alex A. Adjei[14], Cezmi A. Akdis[15], Élie Azoulay[16], Jan Bakker[17,18,19], Zuhair K. Ballas[20], Philip G. Bardin[21], Esther Barreiro[22], Rinaldo Bellomo[23], Jonathan A. Bernstein[24], Vito Brusasco[25], Timothy G. Buchman[26,27,28], Sudhansu Chokroverty[29], Nancy A. Collop[30,31], James D. Crapo[32], Dominic A. Fitzgerald[33], Lauren Hale[34], Nicholas Hart[35], Felix J. Herth[36], Theodore J. Iwashyna[37], Gisli Jenkins[38], Martin Kolb[39], Guy B. Marks[40], Peter Mazzone[41], J. Randall Moorman[42,43,44], Thomas M. Murphy[45], Terry L. Noah[46], Paul Reynolds[47], Dieter Riemann[48], Richard E. Russell[49,50], Aziz Sheikh[51], Giovanni Sotgiu[52], Erik R. Swenson[53], Rhonda Szczesniak[54,55], Ronald Szymusiak[56,57], Jean-Louis Teboul[58], and Jean-Louis Vincent[59]

[1]Department of Medicine and [2]Department of Epidemiology, Columbia University Irving Medical Center, New York, New York; Editor-in-Chief, *Annals of the American Thoracic Society*; [3]Department of Thoracic Medicine, The Prince Charles Hospital, Brisbane, Queensland, Australia; Editor-in-Chief, *Journal of Cystic Fibrosis*; [4]Department of Surgery, University of Cincinnati, Cincinnati, Ohio; Editor-in-Chief, *Respiratory Care*; [5]University of Dundee, Dundee, Scotland; Deputy Chief Editor, *European Respiratory Journal*; [6]London, England; Deputy Editor, *The Lancet Respiratory Medicine*; [7]Department of Medicine, Queen's University, Kingston, Ontario, Canada; Associate Editor for Data Science, *Critical Care Medicine*; [8]Department of Pulmonary Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas; Editor-in-Chief, *Journal of Bronchology and Interventional Pulmonology*; [9]Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland; Deputy Editor-in-Chief, *SLEEP*; [10]Department of Allergy, Kaiser Permanente Medical Center, San Diego, California; Editor-in-Chief, *The Journal of Allergy & Clinical Immunology: In Practice*; [11]Division of Child Health, Obstetrics, and Gynecology, University of Nottingham, Nottingham, England; Joint Editor-in-Chief, *Thorax*; [12]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina; Associate Editor, *Pediatric Pulmonology*; [13]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada; Advisor, *COPD: Journal of Chronic Obstructive Pulmonary Disease*; [14]Department of Oncology, Mayo Clinic, Rochester, Minnesota; Editor-in-Chief, *Journal of Thoracic Oncology*; [15]Swiss Institute of Allergy and Asthma Research, University of Zurich, Davos, Switzerland; Editor-in-Chief, *Allergy*; [16]St. Louis Hospital, University of Paris, Paris, France; Editor-in-Chief, *Intensive Care Medicine*; [17]Department of Medicine, Columbia University Irving Medical Center, and Division of Pulmonary, Critical Care, and Sleep, NYU Langone Health, New York, New York; [18]Department of Intensive Care Adults, Erasmus MC University Medical Center, Rotterdam, the Netherlands; [19]Department of Intensive Care, Pontificia Universidad Católica de Chile, Santiago, Chile; Editor-in-Chief, *Journal of Critical Care*; [20]Department of Internal Medicine, University of Iowa and the Iowa City Veterans Affairs Medical Center, Iowa City, Iowa; Editor-in-Chief, *The Journal of Allergy and Clinical Immunology*; [21]Monash Lung and Sleep, Monash Hospital and University, Melbourne, Victoria, Australia; Co-Editor-in-Chief, *Respirology*; [22]Pulmonology Department, Muscle and Lung Cancer Research Group, Research Institute of Hospital del Mar and Centro de Investigación Biomédica en Red Enfermedades Respiratorias Instituto de Salud Carlos III, Barcelona, Spain; Editor-in-Chief, *Archivos de Bronconeumologia*; [23]Department of Intensive Care Medicine, Austin Hospital and University of Melbourne, Melbourne, Victoria, Australia; Editor-in-Chief, *Critical Care & Resuscitation*; [24]Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio; Editor-in-Chief, *Journal of Asthma*; [25]Department of Internal Medicine, University of Genoa, Genoa, Italy; Editor-in-Chief, *COPD: Journal of Chronic Obstructive Pulmonary Disease*; [26]Department of Surgery, [27]Department of Anesthesiology, and [28]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia; Editor-in-Chief, *Critical Care Medicine*; [29]JFK New Jersey Neuroscience Institute, Hackensack Meridian Health–JFK Medical Center, Edison, New Jersey; Editor-in-Chief, *Sleep Medicine*; [30]Department of Medicine and [31]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia; Editor-in-Chief, *Journal of Clinical Sleep Medicine*; [32]Department of Medicine, National Jewish Hospital, Denver, Colorado; Editor-in-Chief, *Journal of the COPD Foundation*; [33]The Children's Hospital at Westmead, Sydney Medical School, University of

Sydney, Sydney, Australia; Editor in Chief, *Paediatric Respiratory Reviews*; [34]Program in Public Health, Department of Family, Population and Family Medicine, Stony Brook University, Stony Brook, New York; Editor-in-Chief, *Sleep Health*; [35]Lane Fox Respiratory Service, Guy's and St. Thomas' Hospital, London, United Kingdom; Joint Editor-In-Chief, *Thorax*; [36]Department of Pneumology and Critical Care Medicine, University of Heidelberg, Heidelberg, Germany; Editor-in-Chief, *Respiration*; [37]Department of Internal Medicine, University of Michigan, Veterans Affairs Center for Clinical Management Research, Veterans Affairs Ann Arbor Health System, Ann Arbor, Michigan; Innovation Editor, *Annals of the American Thoracic Society*; [38]Department of Experimental Medicine, University of Nottingham, Nottingham, England; Joint Editor-In-Chief, *Thorax*; [39]Department of Medicine, McMaster University, Hamilton, Ontario, Canada; Chief Editor, *European Respiratory Journal*; [40]South Western Sydney Clinical School, University of New South Wales, Sydney, New South Wales, Australia; Joint Editor-in-Chief (Lung Diseases), *International Journal of Tuberculosis and Lung Disease*; [41]Department of Pulmonary Medicine, Cleveland Clinic, Cleveland, Ohio; Incoming Editor-in-Chief, *CHEST*; [42]Department of Medicine, [43]Department of Physiology, and [44]Department of Engineering, University of Virginia, Charlottesville, Virginia; Editor-in-Chief, *Physiological Measurement*; [45]Division of Pulmonary and Sleep Medicine, MassGeneral Hospital for Children, Harvard Medical School, Boston, Massachusetts; Editor-in-Chief, *Pediatric Pulmonology*; [46]Division of Pulmonology, Department of Pediatrics, University of North Carolina, Chapel Hill, North Carolina; Deputy Editor, *Pediatric Pulmonology*; [47]Department of Thoracic Medicine, Royal Adelaide Hospital, University of Adelaide, Adelaide, South Australia, Australia; Co-Editor-in-Chief, *Respirology*; [48]Department of Psychiatry and Psychotherapy, Freiburg University Medical Center, Freiburg, Germany; Editor-in-Chief, *Journal of Sleep Research*; [49]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom; [50]Lymington New Forest Hospital, Hampshire, United Kingdom; Editor-in-Chief, *International Journal of COPD*; [51]Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, Scotland, United Kingdom; Editor-in-Chief: *npj Primary Care Respiratory Medicine*; [52]Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical, and Experimental Sciences, University of Sassari, Sassari, Italy; Associate Editor, *European Respiratory Journal*; Advisor, *COPD: Journal of Chronic Obstructive Pulmonary Disease*; [53]Department of Medicine, University of Washington, Veterans Affairs Puget Sound Health Care System, Seattle, Washington, United States; Editor-in-Chief, *High Altitude Medicine & Biology*; [54]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [55]Department of Pediatrics, University of Cincinnati, Cincinnati, Ohio; Statistical Editor, *Thorax*; [56]Department of Medicine and [57]Department of Neurobiology, David Geffen School of Medicine at UCLA, Los Angeles, California; Editor-in-Chief, *SLEEP*; [58]Centre Hospitalier Universitaire Bicêtre, Le Kremlin-Bicêtre, France; Editor-in-Chief, *Annals of Intensive Care*; and [59]Bicêtre Hospital, Assistance Publique–Hôpitaux de Paris, University Paris South, Le Kremlin-Bicêtre, France; Editor-in-Chief, *Critical Care*

ORCID IDs: 0000-0001-5258-0228 (D.J.L.); 0000-0001-8651-7139 (S.C.B.); 0000-0002-0912-3360 (R.D.B.); 0000-0002-0765-7158 (D.M.M.); 0000-0001-8020-019X (C.A.A.); 0000-0003-2236-7391 (J.B.); 0000-0002-8976-8053 (G.B.M.); 0000-0002-5772-1648 (J.R.M.).

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 nonobese adults? Does caffeine intake protect against pulmonary arterial hypertension? Do antidepressants reduce the risk of the acute respiratory distress syndrome (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
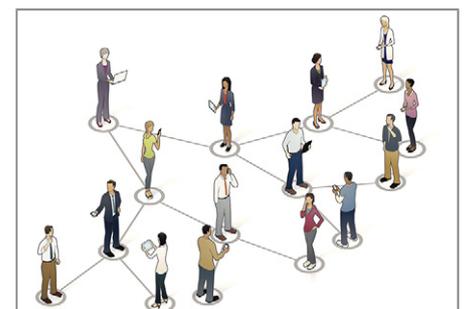
confounding. The latter is addressed in detail later in this document.

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 models.

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. On the basis of our experience, we have identified five 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 on the basis of prior knowledge rather than selecting those variables associated with the exposure or outcome using the available data.

### Using Causal Models to Identify Confounding

Although 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 (on the basis of prior knowledge). Although investigators often feel some discomfort in deciding what causal effects do and do not exist on the basis of 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

**Table 1.** Key principles

---

**Key Principle #1: Causal inference requires careful consideration of confounding**
- Preferred variable selection methods
  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
  3. *P* value– or model-based methods
  4. Methods based on β-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)

**Key Principle #2: Interpretation of results should not rely on the magnitude of *P* values**
- *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.

**Key Principle #3: Results should be presented in a granular and transparent fashion**
- Use the STROBE statement and checklist
- Model tables after the STROBE explanation and elaboration document (30)
- Visual presentation of quantitative results
  ○ 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

---

*Definition of abbreviation*: STROBE = Strengthening the Reporting of Observational Studies in Epidemiology.

**A** *Confounding*

**B** *Mediation*

**C** *Collider Bias*
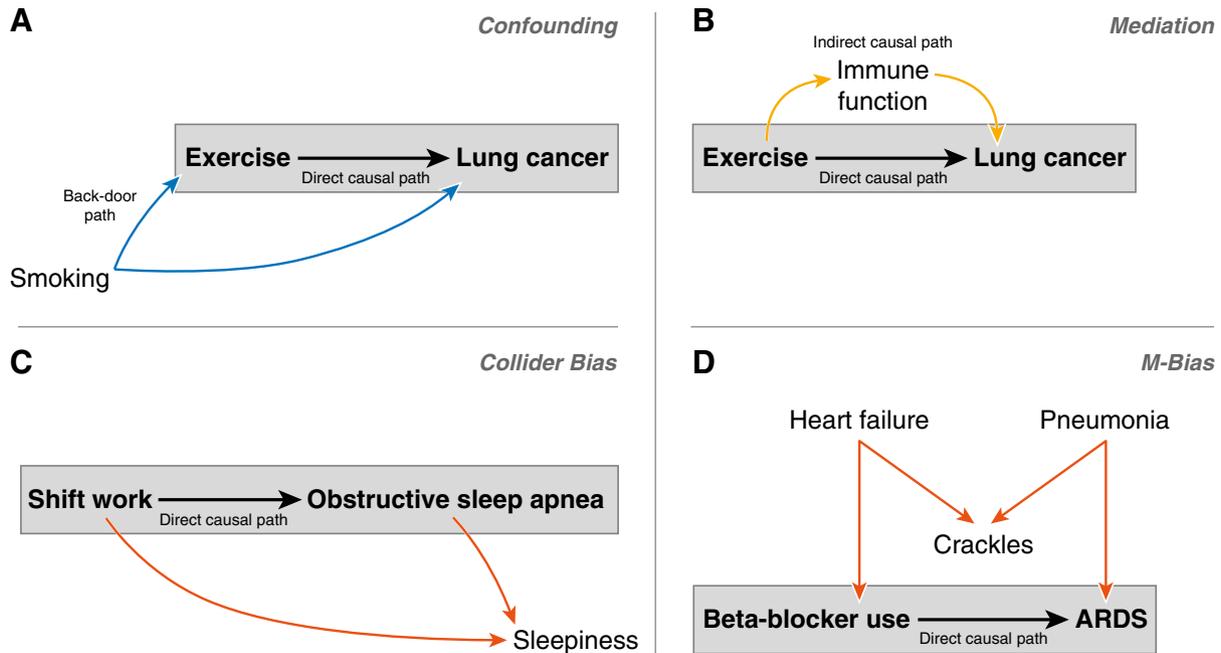
**D** *M-Bias*

**Figure 1.** Directed acyclic graphs illustrating (*A*) confounding, (*B*) mediation, (*C*) collider bias, and (*D*) M-bias. Each arrow represents a causal effect. (*A*) The blue arrows represent an open back-door path: exercise ← smoking → 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. (*B*) The black arrow represents a direct causal path. The yellow arrows represent an indirect causal path. "Immune function" partially mediates the association between exercise and lung cancer: exercise → immune function → lung cancer. Control of "immune function" would be inappropriate, because it would partially close the causal path, attenuating the observed association between exercise and lung cancer. (*C*) The orange arrows represent a closed back-door path: shift work → sleepiness ← 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. (*D*) The orange arrows represent a closed back-door path: chronic β-blocker therapy ← heart failure → crackles ← pneumonia → acute respiratory distress syndrome (ARDS). "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.

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 → lung cancer, and one back-door path: exercise ← smoking → 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, because 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 the immune system: 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 → sleepiness ← 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 β-blocker use and the risk of developing ARDS. We might be tempted to adjust for the presence of auscultatory crackles at hospital admission, because: 1) heart failure leads to both chronic β-blocker therapy and crackles, and 2) 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 β-blocker therapy ← heart failure → crackles ← pneumonia → 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 who 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 nonparametric 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 postpublication 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

**Table 2.** Causal inference resources

**Books**
Pearl J, Mackenzie D. The book of why: the new science of cause and effect. New York, NY: Basic Books; 2018. (17)
Pearl J. Causality: models, reasoning, and inference. New York, NY: Cambridge University Press; 2009. (16)
Hernán MA, Robins JM. Causal Inference. Boca Raton: CRC Press; 2018 (Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/)

**Articles**
Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48. (10)
Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14: 300–306. (9)
Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–625. (11)
Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;20:488–495. (13)
Morabia A. History of the modern epidemiological concept of confounding. *J Epidemiol Community Health* 2011;65:297–300. (12)
Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. *Respirology* 2014;19:303–311. (14)
Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018;108:616–619.

**Websites**
An online course about causal inference and directed acyclic graphs. Hernán M. Causal diagrams: draw your assumptions before your conclusions. By Miguel Hernán. Available from: https://www.edx.org/course/causal-diagrams-draw-assumptions-harvardx-ph559x.
A Web-based environment for creating directed acyclic graphs: http://dagitty.net. 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. (18)

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 (21) and further developed by John Ioannidis (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," because 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, 0.97–4.2), 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 "nonsignificant," 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. ■

**Author disclosures** are available with the text of this article at www.atsjournals.org.

## References

1 Prasad VK, Cifu AS. Ending medical reversal: improving outcomes, saving lives. Baltimore: Johns Hopkins University Press; 2015.
2 Ioannidis JPA. 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;198:859-867.
5 Hernán 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 Hernán 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 Hernán MA, Hernández-Díaz 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: the new science of cause and effect. 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 Inference* 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, *et al*. Redefine statistical significance. *Nat Hum Behav* 2018;2:6–10.
24 Ioannidis JPA. The proposal to lower P value thresholds to .005. *JAMA* 2018;319:1429–1430.
25 Lakens D, Adolfi FG, Albers CJ, Anvari F, Apps MAJ, Argamon SE, *et al*. Justify your alpha. *Nat Hum Behav* 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, *et al*. 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, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. 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, Gøtzsche PC, Mulrow CD, Pocock SJ, *et al*.; STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007;147:W163–W194.
31 Tufte ER. The visual display of quantitative information. Cheshire, CT: 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.