Moving Beyond the Cause Constraint: A Public Health of Consequence, May 2018

See also Hernán, p. 616; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

Both of us have been involved in decades’ worth of population health science writing that has, at times, sidestepped the issue of whether a condition, factor, or circumstance of interest was indeed a “cause,” sometimes nudged by coauthors, other times by editors, others by reviewers, and still others by our own timidity, as we sought to avoid unnecessary argument.

CAUSE AND CONSEQUENCE

It is this timidity that an excellent and provocative commentary in this month’s AJPH by Hernán (p. 616 and p. 625) seeks to dispel. Hernán argues that we should use the word “cause” when we mean cause and persuasively outline the dangers that emerge when we avoid the word “cause.” Several other articles in this issue (see Begg and March [p. 620], Ahern, Chiolero [p. 622], Glymour and Hamad [p. 623], and Jones and Schooling [p. 624]) elaborate on this.

REFERENCES


When we focused on promoting a public health of consequence, it was likely unintentional that our language and thinking returned to the notion of cause in public health; to actually move the needle on population health, we have to move our thinking, our research, and our actions from describing disparities to understanding the causes of them. Reflecting on the arguments introduced by Hernán, we wonder why “cause” has become such
a challenging word to use in our science. We offer here three potential explanations to aid the understanding of what we do and how we do it and, hopefully, to keep our thinking focused on the causes and not find contentment with descriptions and associations.

CAUSE RELUCTANCE

First, our concern with using “cause” emerges from an over-abundance of caution about the limitations of any particular statistical approach to tell us that one cause is the active ingredient that influences the outcome, all other factors being held equal. This concern is tied closely to efforts to isolate causes, replicate assumptions that are achievable only in idealized randomized control trials, and determine what the cause may be so that we may interact on that one cause. This approach, an effort at causal isolationism, has pushed us ever deeper into anxiety about identifying a particular factor as a cause, because we of course recognize that other factors may be equally meritorious causes.

This has implications that extend far beyond the methodological confines of any particular discipline and echo efforts in the broader public discussion about how we act to improve health. Our consideration of causes of death, to provide an obvious example, as being centrally behavioral effectively constrains our spending and commensurate research focus on efforts to modify behavior, neglecting the more upstream social causes of death that are equally responsible for the generation of health and disease.

Loosening the reins on our use of the word “cause,” embodying an appreciation of the multiplicity of causes, stands to create space to introduce the foundational drivers of health in a national health causal conversation, a change that would be welcome indeed.

Second, our reluctance to use the word “cause” has come from the conflation of the act of causal thinking and the application of statistical methods to observed data that collectively have come to be called “causal modeling” approaches. In some respects, this error is not unique to cause and causal thinking. When multilevel modeling emerged as a common useful technique in population health science more than 15 years ago, its widespread adoption also came in some respects at the expense of multilevel thinking. The adoption of statistical approaches that are new and unfamiliar may, reasonably enough, occasion anxiety, perhaps fueled by the relatively small number of scholars who, at first, are familiar and comfortable with the relevant methods. This anxiety about new methods may be unavoidable. It seems incumbent on the early practitioners to do their part to help us not confute unease with statistical methods and causal approaches with a solid understanding of the broader notion of causation. Simply put, an appreciation of causal thinking does not depend on causal modeling, and unfamiliarity with the latter need not scare away engagement with the former.

Third, the reluctance to use the word “cause” represents the long tail diffusion of the science and the attendant loss of nuance that is well accepted among leading thinkers in the field as methods and approaches are adopted by the broader mass of scientists and practitioners. Hernán is saying little new in his editorial; the ideas he brings up have long been recognized and discussed in other articles by leading methodologists, many of whom Hernán cites. The authors of these articles recognize the challenges inherent in causal inference, the potential and limitations of our methods that can help inform our causal thinking. The authors, additionally, realize that this does not mean that we need to throw the cause baby out with the bathwater of overinference from limited studies. However, the “association is not causation” crutch—perhaps emerging as an antidote to overgenerous interpretations of work that we should not hinge much on—has long spread throughout population health science, has been echoed by the lay press, and, alas all too frequently, has been adopted by reviewers. This oversimplification has gone on to chill the use of “cause” by many throughout the field, leading to the challenges that Hernán articulates so well. This calls for ongoing, persistent, and relentless communication by leaders in the field about the nuance that should inform our thinking and a refusal to allow simplistic prohibitions to spread without challenge. That Hernán issues his challenge in the pages of AJPH is laudable; we all probably should have risen to the challenge sooner.

STRIVE FOR CAUSAL THINKING

A recent book defined population health science as “the study of the conditions that shape distributions of health within and across populations, and of the mechanisms through which these conditions manifest as health. An argument that would carry and a desire not to have that word “cause” could substitute for both euphemisms in both places in this definition. But the elision of the word “cause” in this definition was a nod to the intellectual baggage it carries and a desire not to have that intellectual encumbrance become the focal point of arguments about whether these conditions are “causes”—an argument that would have been beside the point of what the definition was trying to achieve to begin with.

We consider the move to reintroduce causal language to population health science welcome and long overdue. The task at hand is to ensure that the language of our writing, and hence of our thinking, accurately reflects what we are trying to do. When we are clear that we are studying causes, we open up the opportunity to identify and act on them. Anything less limits the reach and scope of population health science.
The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Miguel A. Hernán, MD, DrPH

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term “causal” is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results.


You know the story:

Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Many journal editors request authors to avoid causal language, and many observational researchers, trained in a scientific environment that frowns upon causality claims, spontaneously refrain from mentioning the C-word (“causal”) in their work. As a result, “causal effect” and terms with similar meaning (“impact,” “benefit,” etc.) are routinely avoided in scientific publications that describe nonrandomized studies. Instead, we see terms like “association” and others that convey a similar meaning (“correlation,” “pattern,” etc.), or the calculatedly ambiguous “link.”

The proscription against the C-word is harmful to science because causal inference is a core task of science, regardless of whether the study is randomized or nonrandomized. Without being able to make explicit references to causal effects, the goals of many observational studies can only be expressed in a roundabout way. The resulting ambiguity impedes a frank discussion about methodology because the methods used to estimate causal effects are not the same as those used to estimate associations.

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating “causal” as a dirty word that respectable investigators do not say in public or put in print. It is true that observational studies cannot definitely prove causation, but this statement misses the point, as discussed in this commentary.

OF COURSE “ASSOCIATION IS NOT CAUSATION”

Suppose we want to know whether daily drinking of a glass of wine affects the 10-year risk of coronary heart disease. Because there are no randomized trials of long-term alcohol drinking, we analyze observational data by comparing the risk of heart disease across people with different levels of alcohol drinking over 10 years. Say that this analysis yields a risk ratio of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn’t drink wine, for example, they have high enough incomes to buy, besides wine, nutritious food and to take time off to exercise, or have better access to preventive health care.

In other words, the risk ratio of 0.8 may be an unbiased estimate of the ratio of heart disease of 0.8 for one glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.
measure of the association between wine and heart disease, but a biased (confounded) measure of the causal effect of wine on heart disease. Importantly, we knew this before conducting the study. That observational effect estimates may be confounded is not a scientific statement, it is a logical one. The sentence “Your observational effect estimate may be seriously confounded” can never be proven wrong, regardless of how much data are available for confounding adjustment. In fact, the sentence is in the same logical category as “You can die in the next five years.” Sadly, both quoted statements are always logically possible, no matter what data you have at this moment.

We all agree: confounding is always a possibility and therefore association is not necessarily causation. One possible reaction is to completely ditch causal language in observational studies. This reaction, however, does not solve the tension between causation and association; it just sweeps it under the rug.

**CONFLATING THE MEANS AND THE ENDS**

In our example, the driving question for the research is whether modifying wine intake can reduce the incidence of heart disease. That is, the primary scientific aim of this observational study is to quantify “the causal effect of wine on heart disease,” not “the association between wine and heart disease” or (gasp) “the link between wine and heart disease.”

In a parallel universe, we might be able to estimate this causal effect by conducting a randomized trial in which large numbers of people are randomly assigned to different levels of wine intake and forced to comply over 10 years. Fortunately, in our world, such a trial is considered unethical and hence forbidden. Unfortunately, that means that observational analyses become our best chance to quantify the long-term causal effect of wine on chronic diseases. And yet all we can estimate from the data are associations that may not reflect causation. The analysis of the observational study is necessarily associational, even though the goal of the observational study is causal.

Interestingly, the same is true of randomized trials. All we can estimate from randomized trials data are associations; we just feel more confident giving a causal interpretation to the association between treatment assignment and outcome because of the expected lack of confounding that physical randomization entails. However, the association measures from randomized trials cannot be given a free pass. Although randomization eliminates systematic confounding, even a perfect randomized trial only provides probabilistic bounds on “random confounding”—as reflected in the confidence interval of the association measure—and many randomized trials are far from perfect.

Hence we need to use causal language to accurately describe the aims of our research, be it based on randomized trial or observational data. Avoiding the word “causal” in a scientific paper or grant application makes it impossible to express the research aims unambiguously. In the wine example, our goal is to estimate the causal effect of wine on heart disease. Therefore, the term “causal effect” is appropriate in the title and Introduction section of our article when describing our aim, in the Methods section when describing which causal effect we are trying to estimate through an association measure, and in the Discussion section when providing arguments for and against the causal interpretation of our association measure. The only part of the article in which the term “causal effect” has no place is the Results section, which should present the findings without trying to interpret them.

Without causally explicit language, the means and ends of much of observational research get hopelessly conflated. As Rothman put it more than 30 years ago,

> Some scientists are reluctant to speak so bluntly about cause and effect, but in statements of hypothesis and in describing study objectives such boldness serves to keep the real goal firmly in focus and is therefore highly preferable to maddening statements about “association” instead of “causation.”

Carefully distinguishing between causal aims and associational methods is not just a matter of enhancing scientific communication and transparency. Eliminating the causal-associational ambiguity has practical implications for the quality of observational research too.

**BETTER CAUSAL QUESTIONS**

Associational questions are easy to formulate and straightforward to answer when data are available. Are you interested in the association between drinking one glass of wine daily and heart disease at a certain time in a certain population? Just compare the risk of heart disease between individuals who drink one glass of wine daily and those who did not drink wine. The statistical analysis is trivial if the data are accurately measured in the population of interest.

Causal questions, on the other hand, are not always easy to formulate. If we want to estimate the causal effect of drinking one glass of wine daily on heart disease, we first need to explain what we mean by “the causal effect of drinking one glass daily on heart disease.” A helpful approach is to define the causal effect in our population as the causal effect that would have been observed in a hypothetical trial in which individuals in our population had been randomly assigned to either drinking one glass of wine or no wine drinking for some period (say, 10 years). Of course, such a trial is infeasible (and unethical), but that is beside the point. The point is that an observational analysis can be guided by defining the causal effect in a hypothetical trial as the inferential target. In other words, a causal analysis in observational data can be viewed as an attempt to emulate a hypothetical trial—the target trial.

Specifying the target trial is a useful device to sharply define a causal question in an observational analysis, and to better understand the data that are necessary for emulating the target trial. A key advantage of specifying the target trial is that it forces investigators to consider the intervention of interest and the time period during which it takes place. For example, a target trial could assign people to 10 years of daily wine drinking from age 55 years, or to 20 years of wine drinking from age 30 years. Each of these target trials answers a different causal question and therefore, if the trials were actually conducted, they would result in different causal effects. For the same reason,
emulating each of these trials using observational data will require a different set of data on treatments and outcomes. For example, to emulate the second target trial, we will need longitudinal data on wine intake and coronary heart disease since age 30 years. The explicit consideration of the causal goal of the research, via specification of the target trial, facilitates the scientific discussion about data requirements for causal inference.

BETTER CONFOUNDING ADJUSTMENT

If the goal of the observational analysis is causal, adjustment for confounding is generally necessary. In our wine example, the risk ratio of 0.8 may be partly or fully explained by access to preventive care and socioeconomic status, which are correlates of both moderate wine drinking and heart disease. Therefore, if the data analysis does not incorporate adjustment for factors that predict both wine drinking and heart disease, we will suspect that the association measure is confounded and therefore we will be reluctant to interpret it as a causal effect measure.

On the other hand, if the goal of the observational analysis is purely associational, no adjustment for confounding is necessary. Remember, if we just want to quantify the association between wine and heart disease, we simply compute it from the data. If we want to develop a predictive model for heart disease, we include covariates (like wine drinking and number of doctor visits in the last year) that predict heart disease, not only confounders. In associational or predictive models, we do not try to endow the parameter estimates with a causal interpretation because we are not trying to adjust for confounding of the effect of every variable in the model. Confounding is a causal concept that does not apply to associations. There is no such thing as a “spurious association” unless we use the term to mean an association that cannot be causally interpreted—but then the goal of the analysis would be causal, not associational.

By contrast, in a causal analysis, we need to think carefully about what variables can be confounders so that the parameter estimates for treatment or exposure can be causally interpreted. Automatic variable selection procedures may work for prediction, but not necessarily for causal inference. Selection algorithms that do not incorporate sufficient subject-matter knowledge may select variables that introduce bias in the effect estimate, and ignoring the causal structure of the problem may lead to apparent paradoxes. Also, note that the parameters for the confounders cannot be causally interpreted because we do not adjust for the confounders of the confounders, and the adjustment variables may include mediators of the confounder effects, including the treatment itself.

Many readers will correctly point out that there is no guarantee that a causal model incorporates all the confounders and therefore there is no guarantee that the parameter estimate for treatment can be causally interpreted, even approximately. We have gone full circle. Surely there is no guarantee the parameter estimate for treatment can be causally interpreted, but we can have an informed scientific discussion about it only if we have first acknowledged the causal goal of the analysis.

CONCLUSIONS

The lack of clarity regarding the goals of the research has often been justified by the questionable validity of causal inferences from observational data. However, this argument simply conflates the aims and the methods of scientific research. An association measure from an observational analysis may be a biased estimate of a causal effect, but being explicit about the goal of the analysis is a prerequisite for good science. Do we want to estimate the association measure or the causal effect measure? Do we want to determine whether “the sort of people who drink a glass of red wine daily have a lower risk of heart disease” or do we want to determine whether “drinking a glass of red wine daily lowers the risk of heart disease”? Associational inference (prediction) or causal inference (counterfactual prediction)?

The answer to this question has deep implications for (1) how we design the observational analysis to emulate a particular target trial and (2) how we choose confounding adjustment variables. Each causal question corresponds to a different target trial, may require adjustment for a different set of confounders, and is amenable to different types of sensitivity analyses. It then makes sense to publish separate articles for various causal questions based on the same data. By contrast, no target trial or confounder selection is necessary in associational analyses.

Arguably, the biggest disservice of traditional statistics to science was to make “causal” into a dirty word, the C-word that researchers have learned to avoid. Glossing over associational and causal goals in many statistics courses and textbooks has led to widespread confusion among users of statistics. In a perfect example of cognitive dissonance, scientific journals often publish articles that avoid ever mentioning their obviously causal goal. It is time to call things by their name. If your thinking clearly separates association from causation, make sure your writing does too.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant AI02634.

Sander Greenland provided insightful comments to an earlier version of this commentary.

REFERENCES


AJPH PUBLIC HEALTH OF CONSEQUENCE

Cause and Association: Missing the Forest for the Trees

In his commentary, Hernán (p. 616) delineates the dangers of avoiding the term “causal”—the “C-word”—in describing the findings from observational research studies. His message seems to run counter to a central tenet of graduate studies in public health: “association is not causation.” This statement is certainly true and undisputed by Hernán. However, it has become a mantra, at potentially considerable cost, which Hernán addresses in a thoughtful case for resurrecting the C-word in discussing observational study results.

We have considered Hernán’s commentary from the perspective of instructors of an integrated introductory course in epidemiology and biostatistics for master of public health (MPH) students, which we have taught for the past five years. Clearly, such a course must aim to have its students distinguish between association and causation. At first blush, we might reject Hernán’s recommendation. However, it is critical to ensure that students understand the goals of public health research; we agree heartedly that as public health professionals, our goal is to identify not just correlates but actual causes of disease, and to take action. This begs the question: can we accurately convey that although our analysis may be limited to identifying associations, the paramount objective in public health and biomedical research studies is to assess causation? We think the answer is yes.

Our own class centers on assessing potential causes through a hierarchy of study designs that balance rigor and aptness in terwoven with an array of analytic techniques. We point our students to the excellent text by Hulley et al., with chapter 9 focusing on ways to enhance causal inference under observational study designs. Table 9.1 poses five possible explanations for observing an association between an exposure and an outcome in an observational study: chance, bias, effect-cause, confounding, and cause-effect; it thus highlights four explanations in addition to the last explanation, cause-effect, the causal hypothesis being evaluated. Chance refers to random error and the possibility of a spurious association between exposure and outcome. Bias refers to systematic error, also leading to a spurious association. Effect-cause, or reverse causation, underscores for students that an association between exposure and outcome may be real (versus spurious), but opposite to the anticipated direction. Confounding, to which Hernán rightly devotes considerable attention, is also a real association, but not causal with respect to the primary exposure of interest. In our experience, students find this exposition of alternate, noncausal explanations both logical and accessible.

This structure, introduced at the beginning of the course and referenced repeatedly throughout, provides a straightforward way for students to engage with the notions of causation and association, without censorship of the C-word. Following the Hulley model, we describe how statistical methods can be used to reveal an association between two variables in a given data set. We then reinforce the notion that such an association may result from five different scenarios, only one of which is the hypothesis that the selected exposure causes the outcome in the context of a well-operationalized causal question. With four possibilities in addition to cause-and-effect, the course marches through a variety of design and analytic methods that allow us to winnow through the potential noncausal explanations for the observed association. In addition to its intuitive appeal, this approach also provides motivation for the various study designs and analytic methods we wish students to learn.

Although the message of “association is not causation” must remain, we agree that we in academia may have overstated the case, thereby doing a disservice to our students and the field. It is certainly possible, and desirable, to bring discussion of cause back into the literature on observational studies, and it may just lead to better science.

Melissa D. Begg, ScD
Dana March, PhD, MPH

CONTRIBUTORS
Both authors contributed equally to this editorial.

REFERENCES

ABOUT THE AUTHORS
Melissa D. Begg is with the Department of Biostatistics, Mailman School of Public Health, and Office of the Provost, Columbia University, New York, NY. Dana March is with the Department of Epidemiology, Mailman School of Public Health.

Correspondence should be sent to Melissa D. Begg, Professor & Vice Provost, Columbia University, 535 West 116th St, Law Memorial Library MCA4341, New York, NY 10027 (e-mail: mds3@columbia.edu). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This editorial was accepted February 1, 2018.
doi: 10.2105/AJPH.2018.304366
Start With the “C-Word,” Follow the Roadmap for Causal Inference

See also Galea and Vaughan, p. 602; Hernán, p. 616; Begg and March, p. 620; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

In this issue of *AJPH*, Hernán (p. 616) argues that when the aim of an investigation is to estimate a causal effect, researchers should be allowed to say so. I agree, and extend this conversation to argue that if we embrace causal questions we must simultaneously extend this conversation to argue that if we embrace causal questions because it is easy to confuse what you want to do with what you can actually do—estimate a statistical parameter. The beauty of a causal roadmap is that it clarifies which parts of a research process are causal, which parts are statistical, and how they link. There are variations on the roadmap, but they broadly include the following: step 1—articulate the scientific question, including definition of the causal parameter of interest; step 2—link the causal and statistical parameters through assessment of the assumptions under which they are equal (known as identifiability); step 3—estimate the statistical parameter; and step 4—interpret the findings.1–3

In following the roadmap, the causal question starts the investigation, consistent with Hernán’s call. In the subsequent step of identifiability, the specific ways in which a particular application meets or falls short of assumptions required for causal interpretation will become clear. Estimation is a purely statistical step. Finally, interpretation ties the process together by weighing the extent to which assumptions are met (step 2) in considering how strongly to interpret the estimate of the statistical parameter (step 3) with respect to the original causal parameter of interest (step 1).

In this common conflation that the roadmap helps us avoid. A TMLE-based estimate of a clearly defined statistical parameter (step 3) that corresponds to an interesting and plausible intervention on an exposure (step 1), cannot be interpreted as a causal effect if the identifiability assumptions are not met (step 2). By contrast, a conditional odds ratio, estimated with simple logistic regression (step 3) can be interpreted as a causal conditional odds ratio for the effect of the exposure (step 1) if the identifiability assumptions are met (step 2).

Certainly, there are reasonable arguments (with which I would agree) that, as a causal parameter of interest, a conditional causal odds ratio is rarely the most interpretable choice (step 1), but whether it can be interpreted causally is a different issue (step 2). Statistical estimation approaches have different strengths and weaknesses—for example, in terms of flexibility in the parameters estimated and ability to handle time-dependent confounding4,5—and these can be considered (as part of step 3) in deciding which methods are worth considering and, finally, which one is strongest for estimating the statistical parameter of interest.

In sum, I agree that researchers should articulate causal research aims. I would add the critical importance of doing this together with a structured roadmap process that avoids confusion and conflation.

Jennifer Ahern, PhD, MPH

REFERENCES


ABOUT THE AUTHOR

Jennifer Ahern is with the Division of Epidemiology, School of Public Health, University of California, Berkeley.

Correspondence should be sent to Jennifer Ahern, School of Public Health, 50 University Hall, University of California, Berkeley, CA 94720-7360 (e-mail: jahern@berkeley.edu). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This editorial was accepted January 25, 2018. doi: 10.2105/AJPH.2018.304358
Data Are Not Enough—Hurray For Causality!

See also Galea and Vaughan, p. 602; Hernán, p. 616; Begg and March, p. 620; Ahern, p. 621; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

Causal inference is of major importance in epidemiology and public health because the determination that an association between an exposure and a health outcome is causal indicates a potential for intervention to improve this health outcome. In the current issue of *AJPH*, Hernán (p. 616) argues that “using the term ‘causal’ is necessary to improve the quality of observational research.” How can we go, in observational epidemiology, from the usually prudent “association is not causation” to an explicitly assumed causal inference?

I think that epidemiology is primarily a scientific practice to inform public health practitioners, policymakers, clinicians, and citizens to help them make adequate health- and disease-related decisions. The goal is to eventually improve health outcomes. Causality concepts should be in line with this practice, and this is why recent developments in causal thinking in observational epidemiology are inspiring. The key, as argued by Hernán, is first to formulate better causal questions, and second to better adjust for confounding.

Until recently, how to formulate adequate causal questions had neither been formalized nor taught in epidemiology. Indeed, training epidemiologists, health scientists, and public health practitioners in causality has often been limited to the study of complex epistemological and philosophical concepts, but without links to observational research practices (data centered)—beyond the (inadequate) use of Hill causal criteria. The counterfactual, or potential, outcomes approach, with the increasing use of directed acyclic graphs and new statistical tools such as G methods for the analysis of observational data, is a driving force to improve how we formulate causal questions. Such an approach is indeed highly effective for addressing issues of confounding, selection bias, and mediation within the same framework. It is also a powerful tool for differentiating association from causality and intervention in public health. This is why I agree with Hernán’s call to use the “C-word” more explicitly.

However, there are issues to resolve if we want to be sure that using the notion of causality more explicitly will help observational epidemiology and, eventually, inform public health. One issue is the growing use of big data analyses, linked to overconfidence in complex statistical modeling, because more than ever it feeds the confusion between causation, prediction, and association. Applied to the study of multiple and weak associations, with a tolerance for vague and unspecified causal effect, it leads to numerous findings without any relevance for public health. Unfortunately, within data-driven epidemiology, almost all resources are devoted to data management and analyses, leaving no room for causal thinking or for the formulation (before running the analyses) of research questions that are truly answerable by observational epidemiology. Another issue is that most health scientists lack the training to conduct adequate statistical analyses for causal inference. Inadequate and unjustified statistical adjustments are legion in observational studies, leading to false causal claims.

The credibility of observational epidemiology has suffered from too much confidence in data and complex statistical modeling. If done properly, putting causality explicitly at the heart of observational epidemiology will help move this field toward actionable skepticism, that is, move it from a risk factor epidemiology centered on correctly estimating statistical associations to an epidemiology of consequences. Public health needs a data-driven and an evidence-based decision-making model; it has to be explicitly causally justified. Data are not enough—hurray for causality!

*Arnard Chiolero, MD, PhD*

**REFERENCES**


**ABOUT THE AUTHOR**

Arnard Chiolero is with the Institute of Primary Health Care, University of Bern, Bern, Switzerland, and the Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada.

Correspondence should be sent to Arnard Chiolero, Deputy Director and Senior Lecturer, Institute of Primary Health Care (BIHAM), University of Bern, Switzerland (e-mail: achiolero@gmail.com). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This editorial was accepted February 9, 2018. doi: 10.2105/AJPH.2018.304379
Causal Thinking as a Critical Tool for Eliminating Social Inequalities in Health

Health researchers are long overdue to fully embrace causal language, as Hernán (p. 616) recommends. Appropriate analytic methods to answer causal questions differ from methods for noncausal questions (e.g., purely prediction problems), and opaque language hinders appropriate study designs. Although the need for transparency when stating research goals might seem obvious, ambiguous phrasing remains common.

In social epidemiology, fuzzy language is sometimes a defensive tactic to circumvent skeptical reviewers, but such hand waving handicaps the field. Investigators should specify whether the goal is to answer a causal question, while acknowledging that results are only estimates of causal parameters: interpretation is always contingent on additional assumptions, even in randomized controlled studies. This critical distinction between the goals of research (to address causal questions) and the interpretation of results (which correspond with causal parameters under specific assumptions) should underlie the development, implementation, and evaluation of health research.

The causality wars have special problems, and other health inequalities arising from social factors such as education, poverty, and discrimination. It is tempting to assume that when an injustice is obvious, the solutions are equally obvious. Yet, we often falter in translating research into action to ameliorate even the most shocking injustices: seemingly reasonable interventions may not function as anticipated, treatments may affect diverse groups differently, and observational evidence is often too vague to be directly translated to interventions.

Such stumbles highlight the importance of critically considering exchangeability, heterogeneous treatment effects, and consistency in observational research. Although social epidemiologists have many causal questions, we must rely heavily on observational research. Randomization of social factors is usually unethical, infeasible, or both. The methodological framework and tools developed in the causal inference movement of the past 30 years are therefore particularly important for social epidemiologists. Triangulation of results from diverse observational studies is critical.

To understand the designs and analyses that are needed, we must first start by articulating clear causal questions, prior assumptions, and most plausible sources of noncausal association, which might be represented via causal diagrams. Findings from alternative designs, including qualitative research, can often provide complementary evidence to support causal inference.

Methods such as instrumental variables, regression discontinuity, and difference-in-differences approaches often creatively leverage natural variation in exposures arising from quasi-experimental circumstances. These methods often—although not always—address confounding more robustly than conventional analyses of observational data (e.g., regressions with confounder control or propensity score models). Alternative methods can be conceptualized as falling along a spectrum of support for causal inferences, with some designs better accounting for likely biases, recognizing that the “best” design depends on the current state of the evidence and the specific question.

Social epidemiologists may be alienated from causal thinking because of past arguments about whether social constructs, such as race, gender, and inequality, are amenable to rigorous evaluations of causal effects. Such exposures are well within the realm of causal thinking. Approaching the dominant questions in our field with a causal lens is the best strategy to deliver the causal evidence we need to reduce racial, gender, and other health inequalities.

Research on social inequalities is sometimes dismissed as ideologically motivated. Precisely because many people have strongly established opinions about issues such as how poverty affects health, we need to scrupulously apply the most rigorous methods to answer these questions. The next step is a frank discussion about what methods are most likely to deliver causal evidence on the social determinants of health.

M. Maria Glymour, ScD
Rita Hamad, MD, PhD

CONTRIBUTORS
Both authors contributed equally to this editorial.

REFERENCES

ABOUT THE AUTHORS
M. Maria Glymour is with the Department of Epidemiology and Biostatistics, University of California, San Francisco. Rita Hamad is with the Department of Family & Community Medicine and the Institute for Health Policy Studies, University of California, San Francisco. Correspondence should be sent to M. Maria Glymour, ScD, Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158 (e-mail: mglymour@epi.ucsf.edu). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This editorial was accepted February 12, 2018. doi: 10.2105/AJPH.2018.304383
Let’s Require the “T-Word”

In this issue of AJPH, Hernán (p. 616) argues that we should stop avoiding the “C-word”—causality—in articles about observational studies when the research question is a causal question. We agree that authors should clearly specify their purpose in the introduction, including whether the goal is characterization, risk stratification, or assessment of causation, to ensure use of distinct and appropriate statistical model building for descriptive, predictive, or causal questions. However, the interpretation of findings from an observational study assessing relations needs to maintain use of associational language to reduce the likelihood of misinterpretation from the media and the general public. Media coverage, for example, on the benefits of drinking a glass of red wine a day (based on the “French paradox”) resulted in increased red wine sales in the United States in the 1990s. Imagine how much worse this misinterpretation would be if stronger causal language were used in Discussion sections. For red wine and reduced risk of coronary heart disease, a likely explanation for the observed protective association is confounding by higher socioeconomic position, better health status, and greater ability to delay gratification, which enable consumption of one glass of red wine per day and reduced risk of coronary heart disease. Mendelian randomization studies have not found a protective effect of moderate alcohol use on coronary heart disease.

Furthermore, we disagree on many levels with the general notion that imagining an observational study as testing a causal effect in a randomized trial is a useful exercise. It fails to distinguish between the theoretical model and its testing, between an intervention and the mechanism by which it operates, and between the different sources of bias. This type of thinking results in claims that models that use statistical techniques such as inverse probability weights mimic a randomized controlled trial, increasing their use in the literature without clear consideration for best practices. All statistical approaches to analyzing observational data for causal questions assume sufficiently measured and adjusted confounders and predictors of missing data, when historically, many adjusted models from observational studies have identified exposures as beneficial, which were later found to be harmful or to have no effect. A focus on bias from confounding and missing data also may divert attention from pervasive biases that can occur from selection into the study dependent on exposure and outcome. For example, a population representative study inevitably excludes people who have already succumbed to a harmful exposure and who cannot easily be re-created by extrapolating from the survivors, even with the use of inverse probability weighting.

In summary, we agree fully with the importance of being clear about the purpose of a study in the Introduction. However, we do not agree with using language in interpretation of results that suggests that an observational study alone has fulfilled its purpose and correctly identified a causal effect. Moreover, what may be more important than adding the “C-word” to the Introduction is to require authors to add the “T-word”—that is, to explain their underlying theory of causal mechanism, whether it is the underlying biology or the underlying social structures and systems that clarify why the authors hypothesize that exposure x causes outcome y, so that we start off with questions that are most likely to yield effective interventions. Furthermore, requiring an explanation of the causal theory would increase the likelihood of collaboration across disciplines.

Heidi E. Jones, PhD, MPH
C. Mary Schooling, PhD

Contributors
Both authors contributed equally to this editorial.

References

About the Authors
Heidi E. Jones is with the Department of Epidemiology and Biostatistics, City University of New York School of Public Health, New York, NY. C. Mary Schooling is with the Department of Environmental, Occupational and Geospatial Health Sciences, City University of New York School of Public Health. Correspondence should be sent to Heidi E. Jones, PhD, MPH, Associate Professor, Department of Epidemiology and Biostatistics, City University of New York School of Public Health, 53 W 125th St, New York, NY 10027 (e-mail: heidi.jones@ph.cuny.edu). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link. This editorial was accepted February 2, 2018. doi: 10.2105/AJPH.2018.304365
The C-Word: The More We Discuss It, the Less Dirty It Sounds

I thank Chiolero (p. 622), Ahern (p. 621), Glymour and Hamad (p. 623), Jones and Schooling (p. 624), and Begg and March (p. 620) for sharing their reactions to my commentary (p. 625). My impression is that there are few substantial disagreements among us, just differences in emphasis or, in one case, a misunderstanding.

Chiolero and Ahern zero in on a key issue: the need to distinguish the causal question from the procedure used to answer it. As Chiolero puts it, “How to formulate adequate causal questions had [not] been formalized” until recently in health research, and much of the teaching is devoted to “data management and analysis, leaving no room for causal thinking or for the formulation (before running the analyses) of research questions.” Ahern stresses the importance of a structured process, or a roadmap, to ask and answer causal questions using observational data.

Specify the Target Trial

The first step of that process is, in Ahern’s words, “to articulate the scientific question, including definition of the causal parameter of interest.” Glymour and Hamad also highlight the importance of this first step when they state, “We must first start by articulating clear causal questions,” which is especially true in social epidemiology when the goal is translating causal inferences into action.

One way of performing this step precisely is to specify the protocol of the hypothetical randomized trial that would allow us to estimate the causal parameter of interest. We refer to that hypothetical trial as the “target trial.” Some of us have argued that causal questions that cannot be translated into a hypothetical experiment are ill-defined. As a consequence of ill-defined questions, data analyses yield numerical estimates that are not easily interpretable as estimates of causal effect.

Emulate the Target Trial

The second step of the process is to emulate the target trial using a combination of data, empirically verifiable assumptions, and statistical methods. Jones and Schooling are concerned that trying to emulate a target trial may drive too much attention to sophisticated statistical techniques (e.g., inverse probability weighting) for confounding adjustment at the expense of a thoughtful consideration of design issues and of expert knowledge summarized in causal theories. I suppose that the process of specifying and emulating a target trial can be misused but, if that happens, Jones and Schooling will find me by their side fighting for sound design and appropriate incorporation of expert knowledge in the process. Indeed, incorrect causal inferences from observational data are often the result of a flawed emulation of the basic design of the target trial (e.g., choice of time zero and classification of treatment groups) rather than of emulation of its randomized assignment (i.e., insufficient confounding adjustment).

Of course, specifying and emulating the target trial do not imply that our observational study “has fulfilled its purpose and correctly identified a causal effect,” as Jones and Schooling warn us. It just means that (1) we can provide a scientific description of the causal effect that we are estimating, and (2) we have provided our best estimate of that causal effect. But, as Begg and March remind us, even our best estimate may be affected by systematic bias attributable to selection, confounding, or mis-measurement (reverse causation, also cited by Begg and March, can often be viewed as a form of confounding in which an undetected outcome or its precursors confound the effect of treatment on the detected outcome). Because these biases induce associations that do not have causal interpretation, the association estimated from any data analysis is always causally suspect.

Again, the process of specifying and emulating a target trial helps by providing a systematic way to explore each type of bias and its potential influence on the effect estimate. The Cochrane tool has adopted this target trial-based approach to assess the risk of bias of nonrandomized studies.

Triangulate

Ultimately, no single study can produce uncontroversial estimates of causal effect. As Glymour and Hamad point out, some form of “triangulation” of studies will be needed. To quantify a causal effect, triangulation consists in explicitly emulating the target trial of interest using different methods and data sources. When some of those emulations are expected to be differentially affected by bias, investigators can use the imperfect estimates from each emulation to try to pinpoint or bound the magnitude of the true causal effect. The idea is analogous to the process by which travelers obtaining readings of radio waves at different positions can triangulate the position of the radio transmitter.

But the success of triangulation efforts to estimate causal effects requires that “causal” stop being considered the C-word that investigators and editors avoid. Only by precisely defining the causal effect of...
interest will we have a chance of estimating it accurately. In the absence of a precise definition of the causal effect of interest in each study, researchers will end up trying to triangulate study estimates that cannot be triangulated, just as travelers who obtain mixed readings from multiple transmitters cannot locate the position of any of them. Who could blame them for being confused?

Miguel Hernán, MD, DrPH

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant AI102634.

REFERENCES

