Journal of Physiotherapy 66 (2020) 273–277

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Appraisal

Research Note: Causal inference

Health researchers often investigate questions of cause and effect. For example, epidemiologists conduct research designed to understand the causes of specific pathologies, and clinical trialists investigate the effects of health interventions.

While research into cause and effect has always been part of health research, formal methods for making inferences about causal effects[a] – methods for 'causal inference' – are quite new. In fact the mathematical tools that provide a rigorous theoretical basis for causal inference were only developed in the last decades of the 20th century. With these new tools, researchers are now more able to make causal inferences and more able to assess the veracity of causal claims than they were just a few decades ago. The tools of causal inference can be used to support claims about the existence of causal effects and, more importantly, to quantify the magnitude of causal effects.

This Research Note provides an introduction to some of the ideas that underpin modern causal inference, particularly the ideas of Donald Rubin[1–3] and Judea Pearl.[4–8] Future Research Notes will examine specific methods for causal inference.

A causal effect is a difference between potential outcomes

The crucial first step in the development of rigorous methods for causal inference was the formal mathematical definition of a causal effect.

Rubin defined causal effects in terms of potential outcomes. Specifically, he defined the effect of an exposure[b] on an outcome as the difference between outcomes with and without the exposure. For example, the effect of a specific physiotherapy intervention on a person's physical function is the difference between the function the person would have attained if she was exposed to the intervention compared with the function she would have attained if she was not exposed to the intervention. This way of thinking about causal effects is consistent with everyday intuition about cause and effect.

It is only possible to observe, at the level of the individual, at most one of the two potential outcomes that define a particular causal effect. To use the preceding example, if the person actually received the intervention, we could observe her function with intervention but not her function without intervention. If she did not receive the intervention, we could observe her function without intervention but not her function with intervention. So a causal effect is the difference between two potential outcomes, one which happened and may have been observed and the other which did not happen and was not observed. This means, paradoxically, that causal effects are defined by something that did not happen. The potential outcome that did not happen is sometimes called the 'counterfactual' potential outcome.

The inability to observe both the potential outcome with exposure and the potential outcome without exposure presents a problem for researchers who wish to use data to make inferences about causal effects: it means causal effects cannot be directly observed or measured. The inability to directly observe causal effects is called the fundamental problem of causal inference.[9]

The fundamental problem of causal inference might appear to preclude the study of causal effects. However, that is not the case. It is theoretically possible to use observations (data) to estimate the average causal effect of an exposure in a population. An important methodological achievement has been identification of conditions that must be met if researchers are to obtain valid estimates of average causal effects.

Randomisation ensures exchangeability and provides a strong basis for causal inference

In the 1920s, Fisher championed the use of randomisation in experiments.[10] He argued that the only rigorous way to infer causal effects was to conduct experiments in which the units of observation (agricultural plots in Fisher's agricultural research, or humans in medical research) were randomised to experimental conditions.

Decades later, clinical trialists began to use randomisation to determine the causal effects of health interventions.[11] They designed trials in which study participants were randomised to receive or not receive a health intervention. There has been widespread acceptance of the idea that randomisation provides a strong basis for causal inference and that therefore randomised trials are the preferred way to determine the effects of health interventions.

Even though methods for conducting randomised trials had become well established in the 1950s, it was not until 1970s that the role of randomisation came to be understood in terms of potential outcomes.[1] Rubin showed that when representatives of a population are randomised to groups that do and do not receive an intervention, one group provides an estimate of the mean potential outcome that would arise if everyone in the population was to receive the intervention, the other group provides an estimate of the mean potential outcome if everyone in the population was not to receive the intervention, and the difference between (or ratio of) these mean potential outcomes provides an estimate of the average causal effect of the intervention in the population.

Central to this logic is the idea of 'exchangeability'. Groups assembled using randomisation are exchangeable because randomisation ensures that both groups are representative of the same population. Only then does the difference between the mean outcomes of exposed and unexposed groups provide a valid estimate of the average causal effect of exposure. Randomisation enables strong causal inference because it ensures exchangeability of study participants who do and do not receive the intervention.

[a] Here I have followed the cumbersome convention of referring to 'causal effects'. The term 'causal effect' is tautological because all effects have causes. Nonetheless, it is useful to refer to causal effects because it prevents conflation with statistical effects. Statisticians often refer to associations as 'effects' when they mean 'associations', but association is not causation.

[b] The term 'exposure' is used to refer to any variable whose causal effect we are interested in. In health research, the most frequently studied exposures are health interventions (such as exercise) and suspected causes of disease (such as physical inactivity).

https://doi.org/10.1016/j.jphys.2020.07.010

In practice, epidemiologists often want to determine the causal effects of an exposure that cannot be randomly allocated to study participants. A classic example is the causal effects of long-term tobacco smoking: it is not possible to randomise people to groups and then, for many years, make one group smoke and prevent the other group from smoking. When randomisation is not possible it becomes necessary to use data from observational studies to make causal inferences. Like randomised trials, observational studies compare outcomes of exposed and unexposed people. However, unlike randomised trials, observational studies compare people who were allocated to (ie, became members of) exposed and unexposed groups by a non-random process. One of the difficulties of using observational data to make causal inferences is that observational data do not provide the assurance of exchangeability that is provided by randomisation. In the absence of the assurance of exchangeability provided by randomisation, observational studies must try to approximate exchangeability as closely as possible. Part of the challenge of making causal inferences from observational data is to design and analyse the data in a way that makes exposed and unexposed participants effectively exchangeable.

Another way of thinking about causation: a causal effect is the consequence of independently manipulating an exposure

In a preceding section, the causal effect of an exposure was defined as the difference between potential outcomes with and without the exposure. Implicit in this definition is the idea that a causal effect is the change in outcome that would occur if, possibly hypothetically, the level of exposure[c] was changed keeping everything else constant.

To explore this idea, consider a question about the causal effect of body mass index (BMI) on the incidence of knee injury in soccer players. Figure 1 puts this research question into a causal context – the figure is intentionally simplistic, and it is problematic for reasons that will be discussed below, but it suffices to make some key points.

In the figure, the arrows indicate possible causal effects. When we ask 'What is the causal effect of BMI on the incidence of knee injury?' we are implicitly asking something like 'If, hypothetically, BMI could be changed while keeping other factors constant, how much would the incidence of knee injury change?'

It is worthwhile thinking through what is meant by 'keeping other factors constant'. To gain some insight into this, think about the two other factors in the figure that affect the incidence of knee injury: age (which as well as affecting incidence of knee injury could also affect BMI) and agility (which as well as affecting incidence of knee injury could also be affected by BMI). Clearly, if we wanted to know the effect of BMI on incidence of knee injury we would be interested in the effect of changing BMI while keeping age constant. If both BMI and age were allowed to change we could not find out about the pure effect of BMI because then the effects of age would get mixed up with the effects of BMI. On the other hand, if we wanted to know the effect of BMI on incidence of knee injury we would not want agility to be kept constant. That is because agility might be affected by BMI, and it could be that BMI has its effect on incidence of knee injury partly through its effect on agility (that is, agility could be a 'mediator' of the effect of BMI on incidence of knee injury). These considerations suggest that the effect of BMI on incidence of knee injury could be thought of as being the change in incidence of knee injury that would occur if BMI was changed and age was kept constant but agility could be changed by BMI.

This example suggests an alternative way of defining a causal effect. The causal effect of an exposure on an outcome is the change in outcome that would ensue if, possibly hypothetically, the exposure was changed while keeping constant all factors that are not themselves changed by the exposure. This is how Pearl conceptualised causal effects.[6] Pearl's way of thinking about causal effects is mathematically equivalent to Rubin's definition of causal effects based on potential outcomes.[6] One advantage of Pearl's way of thinking about causal effects is that it makes exchangeability more transparent: if we want to find out about the causal effect of BMI on incidence of knee injury we must compare the incidence of knee injury in groups with high and low BMI. The groups will only be exchangeable if they have the same age, but they could be exchangeable even if they do not have the same agility.

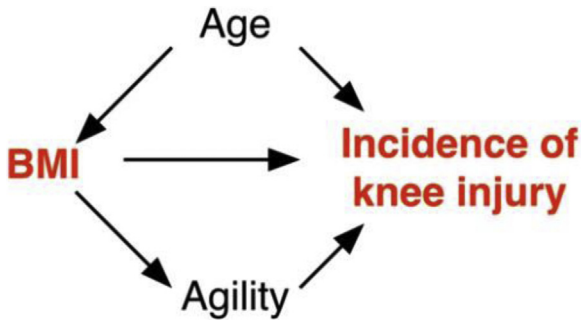

Figure 1. A graphical representation of the presumed causal context for a question about the causal effect of body mass index (BMI) on incidence of knee injuries in soccer players.

One further point needs to be made about the simplistic example used here. The question about whether BMI has a causal effect on knee injury is arguably an ill-specified question that cannot be answered. That is because the causal effect of BMI on incidence of knee injury is the change in outcome that would ensue if BMI was changed, keeping age constant. The problem is that we did not consider how BMI would be changed. BMI could be reduced by exercise, diet or amputation, and not all of these ways of changing BMI would be expected to induce the same change in incidence of knee injury. For that reason, many (not all) epidemiologists would argue that questions about the effects of exposures like BMI (or of many other 'states', including sex and race) cannot be answered unless it is clear how the exposure is to be changed. The interested reader is referred to recent discussions of this issue.[12,13]

Directed acyclic graphs

Figure 1 is a simple example of a 'directed acyclic graph' or DAG. The graph is 'directed' because the arrows are unidirectional (effects must follow causes, so causation can only flow in one direction) and it is 'acyclic' because the arrows do not form a loop (a variable cannot cause itself). DAGs are simple and intuitive – anyone can look at a DAG and see what it means. So it may come as a surprise that DAGs have become amongst the most powerful of all tools for causal inference. Here, the discussion of DAGs is restricted to only as much detail as is required for the following sections.

DAGs provide a simple graphical way of encoding a researcher's beliefs about possible causal relationships between variables. Exactly the same information could be encoded in a set of equations, but DAGs are preferred because they are easier to construct and interpret than equations. When a researcher draws a DAG with an arrow from variable A to variable B, she is articulating the belief that it is possible that A causes B.[d] When she draws a DAG that does not have an arrow between A and B she is asserting that A does not cause B and B does not cause A.[6]

The construction of DAGs must be informed by an understanding of causal relationships. Those understandings may be theoretical (ie, based on presumed mechanisms) or based on data. Sometimes,

[c] Some exposures are binary. An example is having a hip replacement. When exposures are binary, the effect of exposure is the difference in outcomes when exposed and unexposed. Other exposures are quantitative. An example is the distance a person runs every week. When exposures are quantitative, the effect of exposure is the difference in outcomes with different levels of exposure.

[d] Formally, drawing an arrow from A to B asserts that A causes B. But the effect could be trivially small, so drawing an arrow from A to B effectively also allows for A to have no effect on B. Consequently, in practice, the effect of drawing an arrow from A to B is to assert either that A causes B or that A does not have an effect on B. That is not a strong assertion. It is equivalent to asserting only that B does not cause A.

researchers will agree about causal relationships. For example, most researchers would agree that spinal disc herniation could cause back pain and that back pain does not cause disc herniation. So there would be agreement that it would be reasonable to draw a DAG with an arrow going from disc herniation to back pain. Other causal relationships are more contestable – some readers might not agree with the assertion, explicitly made in Figure 1, that agility could affect incidence of knee injury; they might think it more likely that knee injury has an effect on agility. More significantly, some researchers might disagree with the assertion, also explicitly made in Figure 1, that age does not have an effect on agility. To some extent it is possible to use data to test the plausibility of DAGs[14,15] and to repair mis-specified DAGs.[16] Nonetheless, because understandings of the causal relationships between variables are imperfect and the ability to interrogate the veracity of DAGs is limited, DAGs are inevitably provisional, contestable and subject to revision as knowledge evolves.

Association is not causation, but association requires causation

It is well known that association is not causation. It is less well known that variables cannot be associated unless there is some sort of causal link between them. Causal effects are the conduit that links variables and induces associations. This section explains how these two apparently contradictory ideas – 'association is not causation' and 'no association without causation' – can be reconciled.

Consider again the DAG in Figure 1. It includes an arrow from age to BMI, meaning that there could be a causal effect of age on BMI. To the extent that age does have a causal effect on BMI we would expect age to leave an imprint – some sort of statistical trace of its effect – on BMI. That imprint is observable as an association. (A wide range of statistics can be used to quantify associations but in this case the association might be quantified in terms of a correlation between age and BMI, or perhaps a difference between the mean BMIs of young and old people.) The idea here is that the causal effect of one variable on another tends to induce an association between those variables.

Importantly, causal effects can also induce associations between variables that do not have causal effects on each other. To understand how this occurs, look again at Figure 1. As well as allowing for the possibility that age has a causal effect on BMI, the DAG allows for the possibility that age has a causal effect on incidence of knee injury. To the extent that age leaves its imprint on both BMI and incidence of knee injury, BMI and incidence of knee injury have something in common. As a result, there may be an association between BMI and incidence of knee injury. The association between BMI and incidence of knee injury could arise even if there was no causal effect of BMI on knee injury and no causal effect of knee injury on BMI (ie, even if there was no arrow between BMI and knee injury). Thus, the association between BMI and incidence of knee injury potentially provides a biased estimate of the causal effect of BMI on incidence of knee injury. We say that the association between BMI and incidence of knee injury is potentially confounded by age. In general, we expect there to be associations between any pair of variables A and B that have a common cause, even if A does not cause B and B does not cause A.[e]

Remember that our initial question was one about the causal effect of BMI on incidence of knee injury. A naïve researcher might look for evidence of a causal effect by comparing the incidence of knee injury in people with low and high BMI (ie, by looking for an association between BMI and incidence of knee injury). That would be potentially misleading because it is possible that both BMI and incidence of knee injury share a common cause (age). The shared common cause could induce an association between BMI and incidence of injury even if there was no causal effect of BMI on incidence of knee injury. For that reason the association between BMI and incidence of knee injury could provide a biased estimate of the causal effect of BMI on incidence of knee injury.

[e] Associations are induced by common causes but not by common effects. Common effects are called 'colliders'. If A causes B, and C causes B, B is a collider.

If we are to make inferences about the causal effects of non-randomised exposures, we need to extract from the observed association just the part of the association that reflects the causal effect. This is done by adjusting away (or 'controlling for' or 'blocking') the parts of the association that do not reflect a causal effect. In our example, the observed association between BMI and incidence of knee injury needs to be adjusted so that it reflects only the causal effect of BMI on incidence of knee injury (both the direct effect indicated by the horizontal arrow in Figure 1 and the indirect effect mediated by the effect of BMI on agility), leaving out the part of the association that arises because age is a common cause of BMI and incidence of knee injury. In general, the process of making causal inferences from observational data can be seen as a problem of adjusting the association between a putative cause and its putative effect, so that the adjusted association is unconfounded by common causes. If associations can be adjusted so that there is no residual confounding we say there is "conditional exchangeability".

Incidentally, the preceding discussion provides another demonstration of why random allocation of study participants to exposed and unexposed groups allows strong causal inference. When a truly random process is used to allocate participants to exposed and unexposed groups, the only factor that influences whether participants were or were not exposed to the treatment (random allocation) does not also have an independent effect on the trial outcome. So there is no common cause of the exposure and the outcome. As a consequence, there is no confounding and the association between exposure and outcome is purely causal.

Controlling for confounding in observational studies

Researchers have at their disposal a wide range of strategies to help them obtain unconfounded or minimally confounded estimates of causal effects from observational data. Some of those strategies can be incorporated into the study design. Where randomisation to exposed and unexposed groups is not possible, the researcher could choose to include in the study only those people who have a particular level of a variable that is to be controlled. For example, in an observational study of the effect of BMI on incidence of knee injury, the researcher could restrict inclusion to soccer players aged 16 years. Restricting age to a single value prevents any effect of age on either BMI or incidence of knee injury in the study, and if age does not cause BMI or knee injury it cannot cause confounding of the association between BMI and knee injury. Another design strategy is matching. If the study includes, for every person with high BMI in the study, a person with low BMI who has the same ('matched') age, the distributions of ages in exposed and unexposed groups would be the same. In that case (provided that the DAG is properly specified) the groups would be exchangeable and the association between BMI and incidence of knee injury would not be confounded by age.

More often the control of confounding is implemented during the analysis stage, after the data have been collected. Many methods can be used at the analysis stage to control for confounding. These methods differ in the assumptions that they make and the populations that they make inferences about. In the early epidemiological literature, researchers usually sought to control for confounding by conducting stratified analyses.[17] Another time-honoured method, popular in econometrics, has been the use of instrumental variable regression.[18] Perhaps the most widely used analytical method, at least in the epidemiological literature, has been regression adjustment using multivariable linear models. Other approaches include statistical matching methods[19] and matching, stratifying, weighting or adjusting by propensity scores.[20] A particularly powerful approach called doubly robust regression combines both regression adjustment and weighting methods.[21] Another suite of methods is specifically designed for analysing data from longitudinal studies with time-varying confounding affected by past exposure.[22]

All of the analytical procedures used to control for confounding (with the exception of instrumental variable regression) involve, one way or another, fixing, balancing or re-weighting the values of particular variables to block their ability to contribute a non-causal

component to the association of interest. In practice, it is necessary to select which variables to control for. Selection of these variables will usually be the most critical step in the analysis – more critical than the choice of analytical method. Importantly, it is usually not necessary to control for all variables that could cause confounding because controlling for one variable will often block the confounding action of other variables. The decision about which variables to control for should be informed by a DAG. This is discussed in a little more detail in the next section.

Directed acyclic graphs can be used to identify minimal sufficient adjustment sets

The preceding sections discussed the control of confounding in the context of the very simple DAG presented in Figure 1. In that example, control of confounding could be achieved by controlling for just one variable (age). Almost all real-word research using observational data necessitates the construction of more complex DAGs, and more complex DAGs will often represent circumstances in which it is necessary to adjust for more than one variable. Figure 2 is a DAG of typical complexity, showing the presumed causal context for a question about the causal effect of diet on the incidence of pressure ulcers in non-ambulant spinal cord injured patients in Bangladesh.

Imagine we were interested in the question of the effect of diet (perhaps operationalised as caloric intake) on the incidence of pressure ulcers in this population. Our question could be: 'How would the incidence of pressure ulcers change if diet was changed, keeping constant all factors that are not themselves changed by diet?' To answer that question we would want to estimate the association between diet and incidence of pressure ulcers, adjusting the association to control for confounding induced by the common causes of diet and pressure ulcers.

It is not at all easy to see, in this example, which variables cause confounding and which variables would need to be controlled for. Fortunately, there are some simple DAG-based procedures that can be used to identify minimal sufficient adjustment sets. A minimal sufficient adjustment set is a set of variables which, if controlled for (and if the DAG is correctly specified), will remove all confounding. Procedures for identifying minimal sufficient adjustment sets will not be discussed here, but the interested reader can find details elsewhere.[5,7,23] The task of identifying minimal sufficient adjustment sets is made easy with software. Of particular note, DAGitty (available at daggity.net) is free, easy-to-use software that can be used to draw DAGs and identify minimal sufficient adjustment sets.[14] DAGitty shows that, if we wanted to estimate the causal effect of diet on incidence of pressure ulcers, we could adjust the association between diet and incidence of pressure ulcers for the variables in any of five minimal sufficient adjustment sets:

1. Age and Income, or
2. Income and Rural/urban dwelling, or
3. Age, Domestic care, Equipment and Severity of injury, or
4. Age, Domestic care, Equipment and Time spent lying, or
5. Age, Domestic care, Equipment and Time spent sitting.

Provided that the DAG is correctly specified, any of these adjustment sets will do. In practice, smaller adjustments sets are preferred over larger adjustment sets, and adjustment sets that include variables measured with little error are preferred over adjustments sets that include variables that are hard to measure. In this example the preference might be to adjust for age and income.

This approach to the task of selecting variables to adjust for differs from the approach most frequently used by epidemiologists. Historically, epidemiologists have tried to identify and adjust for all potential confounders.[f] That approach is suboptimal for several reasons,

[f]The definition of confounders is problematic.[27] It is much easier to identify confounding than to define confounders.

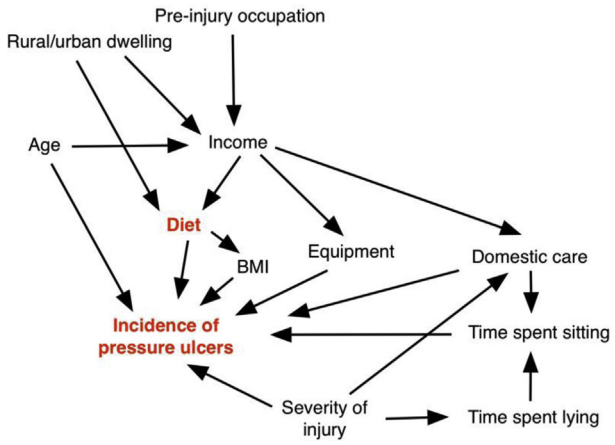

Figure 2. A directed acyclic graph (DAG) showing the presumed causal context for a question about the causal effect of diet on the incidence of pressure ulcers in non-ambulant spinal cord injured patients in Bangladesh.

including that it requires the collection of more data than is necessary, involves estimation of more statistical parameters than is necessary, unnecessarily complicates the analysis, encourages inclusion of poorly measured variables, and risks adjusting for variables (mediators or colliders) that increase rather than decrease bias. A better approach is to draw a DAG and use the DAG to select a minimal sufficient adjustment set.

Concluding remarks

Researchers who wish to explore causal questions should, wherever possible, conduct randomised studies. It is possible to conduct randomised studies of most health interventions, so the effects of health interventions should usually be studied with randomised trials. On the other hand, it is usually not possible to conduct randomised studies of factors thought to cause disease. Consequently studies of disease aetiology will usually need to be observational studies.

Randomisation goes a long way to enabling robust causal inference. But causal inference can be difficult, even in randomised trials, if the trial is imperfect (eg, if there is substantial loss to follow-up, high levels of non-compliance or unplanned treatment crossover). Methods developed for observational studies can sometimes be used to salvage estimates of causal effects from imperfect randomised trials.[24]

Obtaining unbiased estimates of causal effects from observational data is more difficult than obtaining unbiased estimates of causal effects from randomised studies. Theoretically, it is possible to obtain unbiased estimates of causal effects from observational data if the analysis is informed by a DAG that is complete and if the statistical model used to analyse the data is properly specified (eg, if assumptions about distributions of observations, linearity of effects and interactions between variables are satisfied). Inevitably there will be uncertainty about the completeness of the DAG and the structure of the statistical models, so estimates of causal effects obtained from observational studies will always be subject to some doubt. However, it is possible to conduct sensitivity analyses to determine whether it is plausible that a study's primary findings could be substantially biased by residual confounding.[25] This should be standard practice in all observational analyses of causal effects.

Many researchers respond to the difficulty of making causal inferences from observational data by being careful never to admit to an interest in causal hypotheses. Instead of framing their research questions in terms of causal effects, they use ambiguous terminology to obscure the causal objective. For example, many researchers state that the objective of their research is to 'explore associations' or

'identify risk factors'. (What is a risk factor? Is it a *causal* factor? That is usually not made clear.) When reporting their findings, some researchers fastidiously avoid the use of the 'c' word.[26] That is very problematic. There is little intrinsic value in knowing about associations or risk factors, so exploration of associations and identification of risk factors is not an intrinsically worthwhile research goal. Researchers who are interested in causal effects should make their interest in causal effects explicit in the statement of their research objectives, and they should explicitly consider the strength of the causal inferences that can be drawn from their data when discussing their research findings.

Competing interests: Nil.

Sources of support: Nil.

Acknowledgments: Rob Herbert is supported by a research fellowship from the Australian NHMRC.

Provenance: Invited. Not peer reviewed.

Correspondence: Robert D Herbert, Neuroscience Research Australia (NeuRA) and School of Medical Sciences, University of New South Wales, Australia. Email: r.herbert@neura.edu.au

Robert D Herbert

Neuroscience Research Australia (NeuRA) and University of New South Wales, Australia

References

1. Rubin DB. *J Educ Psychol*. 1974;66:688–701.
2. Imbens GW, Rubin DB. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press; 2015.
3. Little RJ, Rubin DB. *Annu Rev Public Health*. 2000;21:121–145.
4. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York: Cambridge University Press; 2009.
5. Greenland S, et al. *Epidemiology*. 1999;10:37–48.
6. Pearl J. *Stat Surv*. 2009;3:96–146.
7. Pearl J, et al. *Causal Inference in Statistics. A Primer*. Chichester: Wiley; 2016.
8. Pearl J, McKenzie D. *The Book of Why. The New Science of Cause and Effect*. New York: Basic Books; 2018.
9. Holland PW. *J Am Stat Assoc*. 1986;81:945–960.
10. Fisher RA. *Statistical Methods for Research Workers*. 5th ed. Edinburgh: Oliver and Boyd; 1934.
11. Bradford Hill A. *Principles of Medical Statistics*. 7th ed. London: The Lancet; 1961.
12. Hernan MA, Taubman SL. *Int J Obes*. 2008;32(Suppl 3):S8–S14.
13. VanderWeele TJ. *Epidemiology*. 2018;29:e24–e25.
14. Textor J, et al. *Int J Epidemiol*. 2016;45:1887–1894.
15. Bollen KA, Pearl J. Eight myths about causality and structural equation models. In: Morgan SL, ed. *Handbook of Causal Analysis for Social Research*. Dordrecht: Springer; 2013.
16. Oates CJ, et al. *Epidemiology*. 2017;28:548–552.
17. Rothman KJ, et al. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott-Raven; 2008.
18. Angrist JD, et al. *J Am Stat Assoc*. 1996;91:444–455.
19. Stuart EA. *Stat Sci*. 2010;25:1–21.
20. Austin PC. *Multivar Behav Res*. 2011;46:399–424.
21. Bang H, Robins JM. *Biometrics*. 2005;61:962–972.
22. Mansournia MA, et al. *BMJ*. 2017;359:j4587.
23. Jewell NP. *Statistics for Epidemiology*. Boca Raton: Chapman & Hall/CRC; 2004.
24. Herbert RD, et al. *BMC Med Res Methodol*. 2018;18:48.
25. VanderWeele TJ, Ding P. *Ann Intern Med*. 2017;167:268–274.
26. Hernan MA. *Am J Public Health*. 2018;108:616–619.
27. VanderWeele TJ. *Ann Stat*. 2013;41:196–220.