

Causal diagrams, information bias, and thought bias

Eyal Shaha¹
Doron J Shaha²

¹Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health;

²Departments of Physics and Mathematics, College of Science, The University of Arizona, Tucson, AZ, USA

Abstract: Information bias might be present in any study, including randomized trials, because the values of variables of interest are unknown, and researchers have to rely on substitute variables, the values of which provide information on the unknown true values. We used causal directed acyclic graphs to extend previous work on information bias. First, we show that measurement is a complex causal process that has two components, ie, imprinting and synthesizing. Second, we explain how the unknown values of a variable may be imputed from other variables, and present examples of valid and invalid substitutions for a variable of interest. Finally, and most importantly, we describe a previously unrecognized bias, which may be viewed as antithetical to information bias. This bias arises whenever a variable does not exist in the physical world, yet researchers obtain “information” on its nonexistent values and estimate nonexistent causal parameters. According to our thesis, the scientific literature contains many articles that are affected by such bias.

Keywords: causal diagrams, derived variables, directed acyclic graphs, imputation, information bias, thought bias

Introduction

In science, bias is the difference between the expected value of an estimator, ie, the process that generated an estimate, and the truth. The truth we seek is a causal parameter which is assumed to exist, eg, the hazard ratio of breast cancer for the contrast between two genotypes. On the ratio scale, the causal parameter may be any positive value, including 1 (no effect).

The truth remains unknown, at least in part, because the values of variables of interest are unknown, and they are forever missing. All that we can do in science is to replace the unknown true values with the information we have obtained about them. That process may generically be called imputation, because imputation in statistics means replacing a missing value. To emphasize the difference between a variable of interest, the values of which are unknown (say, A) and its imputation, we will denote the latter variable by the subscript I , and use the symbolic expression “ $A \Theta A_I$ ” (read “ A is imputed by A_I ”). For instance, if A denotes the genotype of participants in a study of genotype and breast cancer, then A_I may denote the genotype variable that was entered into a regression model. In an indeterministic world, it is always possible for the value of A_I to be different from A , because A_I is still undetermined when A has already taken a value. Therefore, information bias, whenever it exists, cannot be eliminated, and can only be reduced.

Correspondence: Eyal Shaha
Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, The University of Arizona, 1295 N. Martin Ave, Tucson, AZ 85724, USA
Tel +1 520 626 8025
Fax +1 520 626 2767
Email shaha@email.arizona.edu

In this article, we use causal directed acyclic graphs (DAGs)¹ to extend previous work on information bias.²⁻⁴ First, we propose a strict definition of the term “measurement” (which differs from the informal use of this word), and show that most imputations by measurement alone ($A \Theta A_i$) are actually a combination of measurements, as redefined, and synthesized information. Second, we explain how the values of A may be imputed from other variables in a two-step process, ie, theory-driven derivation of a substitute variable followed by imputation of its values. Third, we show examples of valid and invalid substitutions for A . Finally and most important, we describe a previously unrecognized bias, which may be viewed as antithetical to information bias. The bias arises whenever A does not exist, and therefore, the so-called information on A is useless and misleading.

Scientific ideas are developed over time as new questions are contemplated, mistakes are corrected, and deeper insight is gained. This article is no exception. Its origin may be traced to our previous analyses of the merit of body mass index,⁵ the so-called metabolic syndrome,⁶ and change variables.⁷ Here, we develop generic ideas that go beyond specific variables and are relevant to all branches of science.

Variables and three worlds

The scientific process begins by specifying the variables in question such that their content may be grasped clearly and intersubjectively. Body fat, height, velocity, and genotype fit this criterion, for example, whereas stamina and athleticism do not. Stamina and athleticism are, undoubtedly, ideas that people hold, but not every idea necessarily holds the status of a variable in the physical world.

Our understanding of a variable is usually a combination of senses and theories. Our senses, along with causal theories, make us claim that velocity, position, mass, genotype, and blood volume are natural variables, that is, they are properties of physical objects. In contrast, variables such as body mass index, change in cognitive function, left ventricular ejection fraction, and metabolic syndrome status are derived variables.⁵⁻⁷ They exist because mathematical ideas exist; they would have not existed in a world with no mathematics. Following Popper’s writing,⁸ we may say that natural variables belong to World 1, the physical world, whereas derived variables belong to World 3, the autonomous world of ideas, which includes mathematics. World 2, in Popper’s classification, contains mental and psychological states.⁸ More generally, it contains the ideas of World 3 when they exist in some minds.

The causal structure of World 1 is the subject of the natural sciences, where causation operates in one direction along the time axis and all variables are time-indexed. In contrast, World 3 causation operates in a radically different way, in that cause and effect may be reversed and time does not play any role. For example, in World 3, the function $Y = 2X$ places X as a cause of Y , but it may be reorganized to place Y as a cause of X : $X = Y/2$. Both versions coexist, and neither version requires any time order for X and Y . Popper also eloquently described how World 1 and World 3 interact via World 2.⁸ World 3 and its connection to World 2 are essential for the method of science, but its peculiar internal causation is not the subject matter of science. Although World 2 variables (eg, psychological states) are relevant to medical research, most of our discussion will focus on natural variables (World 1).

Imputing a variable by “measurement alone”

Not all natural variables are measurable by contemporary technology, but some of them are. However, what is meant by the phrase “measurement of a variable”? Although common usage invokes the image of a human observer or a measuring device, we propose a far more restricted idea, ie, that a variable is “measured” whenever its value is imprinted (copied) onto another location. Any imprinting of the value of one variable onto another is a measurement, an attempt to copy the information directly. For example, a plain stadiometer measures the height of a person, because the height is imprinted on the ruler, independent of any observer. Similarly, a fingerprint measures the morphology of the epidermal ridges, and a chest X-ray measures the cardiac silhouette.

Measurement in that strict sense is rarely, if ever, the only step in research. In the case of a stadiometer, light from the ruler reaches the observer’s retinas, and a complex causal process leads to the storing of a number (and units) in the observer’s brain, say, 5 ft. Then, other directed causal paths may transfer “5 ft” from the observer’s brain to a computer file via hand motion, writing, reading, keying, and the electrical signals of a computer. Many of the variables in these causal paths are not the measured version of a previous variable (in the sense of simple imprinting), but are synthesized from previous information, sometimes on a different scale. In the previous example, “5 ft” on a stadiometer is synthesized in the observer’s brain from electrical signals in the optic nerves, a causal process that cannot be summarized as “value-to-value imprinting”. Similarly, “5 ft” (or even

“1.52 m”) on a computer screen is synthesized from a set of binary variables.

Consider another example, ie, “measuring” the use of a sleeping pill. The value of the binary variable “sleeping pill swallowed” is transferred to the participant’s brain by synthesizing information from various body sites, eg, the eyes, hands, tongue, and throat. That value may be transferred to the participant’s vocal cords in response to the question “did you swallow a sleeping pill last night?”, from which it may be transferred to the interviewer’s eardrums, and so on. Again, many of the steps are not measurements; they make up cause-and-effect relationships, but are not the simple copying of the value of one variable onto the next.

Two components of imputation, namely imprinting and synthesizing, are illustrated in Figure 1, where an asterisk indicates intermediary A variables between A and A_I . The variable A^* is copied onto A^{*2} , whereas A^{*3} is obtained by synthesizing information about A^{*2} from the variables B , C , D , and E . Accordingly, there are two structural sources for the failure of A_I (the imputed version of A) to be identical to A . First, a natural variable external to the A sequence (denoted by M) causes inaccurate imprinting of a value of A^* onto A^{*2} . For instance, the participant’s hearing status (M) affects his response (A^{*2}) to a question about taking a sleeping pill. Second, imperfect synthesis of A^{*3} from B , C , D , and E would cause A^{*3} to be different from A^{*2} . For instance, the participant’s response, as perceived by the interviewer (A^{*3}), might not be synthesized accurately from the sounds that reached the interviewer’s eardrums (B , C , D , and E).

In summary, what has traditionally been called measurement of A is better described as the transfer of information from A to A_I along directed causal paths ($A \rightarrow \dots \rightarrow A_I$). Some

of the steps take the form of simple imprinting of the previous value, whereas many steps involve synthesizing information from other variables. The transferred information is subject to interference by external variables on the one hand, and by inaccurate synthesis of information on the other.

In some cases, A_I provides information about the values of A because the two variables are associated via a prior A , which is their common cause: $A_I \leftarrow A_{PRIOR} \rightarrow A$. Such a causal structure could operate in an intention-to-treat analysis of a randomized trial,⁹ where A_{PRIOR} is the randomized allocation, A is the treatment offered ($A_{PRIOR} \rightarrow A$), and A_I is an effect of A_{PRIOR} through another path of transferred information ($A_{PRIOR} \rightarrow \dots \rightarrow A_I$).

Imputing a variable from other variables

In many circumstances, A cannot be imputed by a combination of imprinting and synthesizing information. Velocity, for instance, cannot be imprinted in physics experiments, and the vital status of myocardial cells cannot be imprinted in observational research. In other circumstances, the physical constraints of a particular study require an alternative method for the imputation of A . When that is the case, A_I may be derived from other variables. To explain the rationale and the process, we first describe the basic features of variable derivation.

Although not widely appreciated, deriving a variable is a causal process in which the derived variable, D , is caused by the variable(s) from which it was derived.⁷ But unlike natural causation, theoretical derivation operates in the world of ideas (World 3), where the universe of functions, denoted F , serves as a causal agent. Figure 2 (DAG A) shows the generic structure: the causes of D are F (whose values

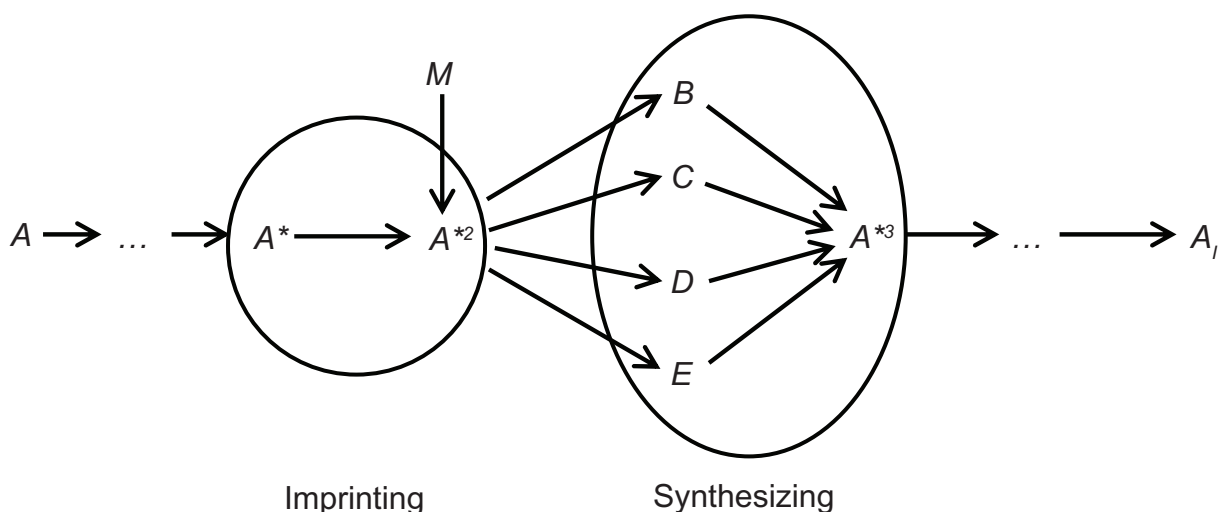


Figure 1 The causal structure of imputation by “measurement alone”.

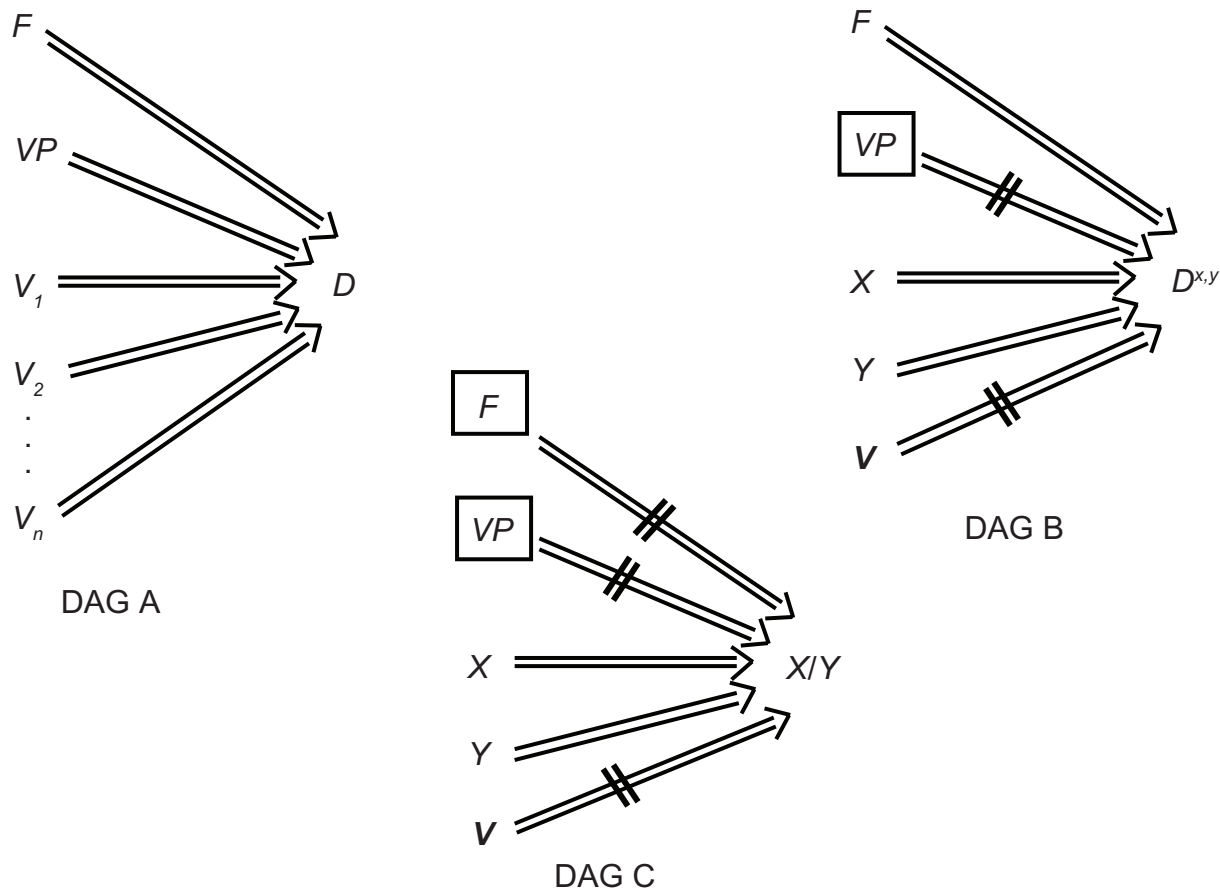


Figure 2 The causal structure of variable derivation: any derivation (directed acyclic graph A); by any function of two variables (directed acyclic graph B); by a specific function of two variables (directed acyclic graph C).

are functions); VP (whose values are variable placement in a function of F), and the variables V_1, V_2, \dots, V_n . DAG B in Figure 2 shows the causal structure whenever D (renamed $D^{x,y}$) is derived from any function of only two variables, X and Y (eg, $X/Y, X^2Y, X-Y$). That derivation implies conditioning on VP (denoted by a box), which dissociated VP and all other variables (denoted V), except X and Y , from D (denoted by two lines over an arrow). DAG C shows the structure for derivation by one specific function of X and Y (X/Y). In this case D is renamed X/Y . To emphasize that the causal process operates in World 3 (with no room for indeterminism), the arrows are drawn in the logic notation of “implied”. In the interest of simplicity, we will later omit VP from the DAGs and assume that VP is restricted to those variables from which D is derived.

Figure 2 follows our notation for variables that contain unknown true values. Therefore, the values of D , just like the values of its makers, are permanently missing. We refer to D as a theoretically derived variable, distinguishing it from its imputed companion, D_I . In practice, D_I is usually

obtained from the imputed version of the makers of D (eg, X_I and Y_I), by first computing D^* , eg, $D^* = X_I/Y_I$. Then, the information is transferred from D^* to D_I along directed paths ($D^* \rightarrow \dots \rightarrow D_I$).

Derivation is mathematical (logical) causation with almost no limits. We can select any set of variables, choose a function, and derive D . Some derivations, however, are not merely a mathematical manipulation of variables,^{5,6,10,11} or equivalent representation of the original information.¹² They serve to impute the values of immeasurable, or unmeasured, natural variables. When that is the case, the derivation is founded not only on mathematics, but also on a theory for why the (unknown) values of D (a derived variable) substitute for the (unknown) values of A , a natural variable. Such a D will be renamed A_S (short for $A_{SUBSTITUTE}$). Then, the known values of D_I will assume the role of A_I .

It is crucial to understand that the process requires two steps. First, we assume that some derived variable, D , provides information on the values of a natural variable,

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021 For personal use only.

A (and therefore label it A_s). Second, we impute the unknown values of that D ($D \ominus D_l$). If the first step is valid, D_l may be renamed A_l because its values also serve to impute the values of A itself. We will argue later that if the first step is not valid (D does not substitute for a natural variable), imputing the values of D is useless.

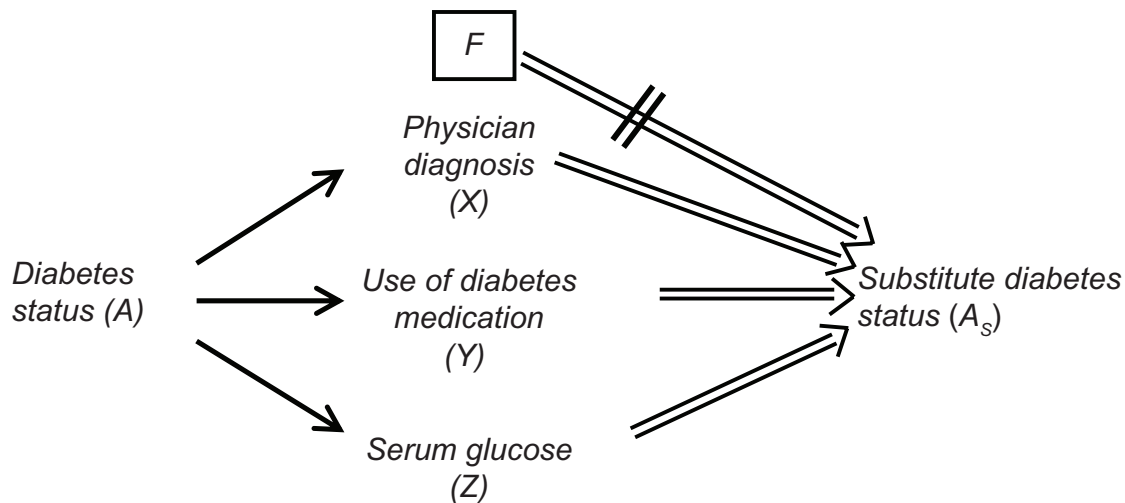
Figure 3 (DAG A) shows an example where the natural variable, A , is diabetes status, ie, a binary variable. We should first ask whether A is indeed a natural variable. Does it exist in the physical world? Is there a sharp demarcation between having diabetes and not having diabetes, or does the underlying natural variable take on a continuous form? Although the answer might not favor a binary variable,¹³ we can assume that the variable exists, for didactic reasons.

If A exists in a binary form, it currently cannot be measured, because there is no known measurement process for a natural variable called diabetes status. Nonetheless, a binary variable may be derived from the reported diagnosis of diabetes (X), use of a diabetes medication (Y), and serum

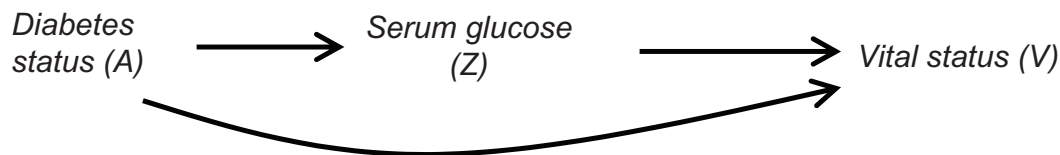
concentration of glucose (Z). That derived variable may be labeled A_s because we hold theories according to which all three variables are effects of A , and have chosen a specific function (Figure 3, DAG A). The causal connections $A \rightarrow X$, $A \rightarrow Y$, and $A \rightarrow Z$ imply associations with A , which means that the values of A may be predicted from the values of X , Y , and Z . But which prediction model should be used? To which value should we restrict the variable F , the universe of all functions?

Numerous functions may be proposed, and there is no means of knowing the best choice. Researchers usually describe their choice in words, rather than in function notation. Consider, for example, the following (partial) quote:

“Baseline diabetes was defined as either a self-reported physician’s diagnosis of diabetes, (or) use of hypoglycemic medications, (or) non-fasting serum glucose levels greater than 200 mg/dL ...”¹⁴



DAG A



DAG B

Figure 3 Theory-based substitution of derived diabetes status (A_s) for unknown, true diabetes status (directed acyclic graph A); causal theories about the effect of diabetes on vital status (directed acyclic graph B).

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021 For personal use only.

If we denote physician diagnosis by X , the use of a diabetes medication by Y , and serum glucose concentration by Z , then the text above corresponds to the following value of F (using 0 and 1 as the values of binary variables):

The value of F is the function:

$$\begin{cases} 1, & \text{if } X=1 \text{ or } Y=1 \text{ or } Z>200 \\ 0, & \text{otherwise} \end{cases}$$

And in practice, $A_s \ominus A_I$ as follows:

$$A_s \ominus A_I = \begin{cases} 1, & \text{if } X_I=1 \text{ or } Y_I=1 \text{ or } Z_I>200 \\ 0, & \text{otherwise} \end{cases}$$

Many researchers refer to this process as a “definition”, and they are often mistaken. First, a definition is no more than a shorthand version for a longer phrase, which makes communication shorter and easier. Life without a definition might be more cumbersome, but no scientific knowledge would be missed.¹⁵ For example, our revised definition of “measurement” might be helpful, but it is not new scientific knowledge. Second, many so-called definitions, including the diabetes definition above, are actually substitution functions for the true values of a natural variable, which is assumed to exist. They are not introduced for the sake of simplified communication. Most important, as we explain later, derived variables, the derivations of which are not founded on a substitution theory (D substitutes for A), are useless in science. They do not generate the causal knowledge they are thought to generate.

The derivation of A_s requires as few as one variable, X , which is associated with A , and a function. The causal structure behind the association between A and X may take one of several forms, ie, cause-and-effect ($A \rightarrow X$ or $X \rightarrow A$), a common cause ($A \leftarrow C \rightarrow X$), or even conditioning on a common effect ($A \rightarrow C \leftarrow X$). Sometimes X may be the variable itself at a later time (A_2), substituting for A at an earlier time (A_1), because $A_1 \rightarrow A_2$. In many substitutions two or more variables are used, but again, each variable is assumed to have one of the above relations with A . The function is usually chosen by one of three methods, ie, a mathematical constraint, an explicit theory, or an empirical method.

If we wish to impute diabetes status (binary) from fasting serum glucose (continuous), a cutoff (step) function is a mathematical necessity. There is no other choice besides the choice of the cutoff point(s). If we wish to impute left ventricular stroke volume (SV) from end diastolic volume (EDV) and end systolic volume (ESV), we invoke the theory

of conservation of matter: the blood that disappeared from the left ventricle must have been ejected. Therefore, $SV_s \ominus SV_I = EDV_I - ESV_I$. An empirical function is usually used when A_I was obtained by measurement and synthesis for some subjects, but is missing for other subjects. In such cases, the nonmissing part of the sample may be used to develop a prediction equation for A_s from a set of variables that are associated with A . For instance,

$$A_s \ominus A_I = \beta_0 + \beta_1 X_I + \beta_2 Y_I + \beta_3 Z_I$$

Eventually, some of the values of A_I are obtained by “measurement”, whereas others are imputed from other variables by an empirical function.

Invalid substitutions of a variable

We describe here several examples, in the context of causal inquiry, where it would be a mistake to derive a substitute variable, A_s , and use its imputed values as the imputed values of A . Such invalid substitutions belong to the category of information bias. To simplify, we discuss derivations from just one variable, labeled Z , using an unspecified function of Z .

First, it is not valid to derive A_s from Z , if we assume that A is not associated with Z . For instance, it is not valid to derive stroke volume from blood type if the two variables are not associated by some mechanism, because blood type does not contain any information on the value of stroke volume.

Second, it would be invalid to derive A_s from Z if we wish to estimate the effect of A on Z or the effect of Z on A . For example, if missing values of carotid wall thickness (A) were empirically imputed from weight (Z), the effect of weight on carotid wall thickness should not be estimated. The reason is simple: the association between these two variables already contains an imputation-driven component. The effect of Z on A cannot be estimated unbiasedly after the imputation.

Third, Z should not be used to derive A_s if our causal theories require conditioning on Z to estimate the effect of A on some variable, V . A variable cannot play a dual role; either it is used to derive A_s , or it is used to estimate the conditional association of A with V .

Why is it invalid to derive A_s from Z when we have to condition on Z ?

It is easy to recognize the problem from the perspective of conditioning by restriction, without losing generality. Suppose that Z serves to derive A_s according to some function $A_s = f(Z)$. Restricting Z , however, to $Z = z$ fixes the value of A_s to $f(z)$, thus, preventing us from using A_s to estimate the effect of A on V .

There are three common reasons for conditioning on Z , ie, to explore effect modification by Z , to block a directed path via Z , such as $A \rightarrow Z \rightarrow V$ in order to estimate the effect of A on V by other paths, and to block a confounding path such as $A \leftarrow Z \rightarrow V$. In all three circumstances, it would be invalid to derive A_s from Z (although sometimes we may be able to keep the derivation and find an alternative for conditioning on Z).

Figure 3 (DAG B) shows two causal theories. According to the diagram, diabetes status affects vital status via serum glucose as an intermediary and via another “direct” causal path. Using serum glucose to derive diabetes status (A_s) is valid, if the goal is to estimate the overall effect of diabetes on survival by all paths. If, however, we wish to estimate the “direct” effect of diabetes on survival, we should condition on serum glucose to block the path “diabetes status (A) \rightarrow serum glucose (Z) \rightarrow vital status (V)”. But in that case serum glucose should not be used in the imputation of diabetes status (as was done in DAG A).

Another common reason for conditioning on Z is confounding. Z might be a confounder for the effect of interest or might reside on a confounding path. Figure 4 shows two

examples where A_s is derived from Z , yet conditioning on Z might be required to estimate the effect of A on V . We used the same notation as in Figure 2, ie, a box around a variable depicts conditioning, and two lines crossing an arrow denote a disrupted association (due to conditioning). A superimposed X indicates an invalid substitution of A_s for A .

DAG A in Figure 4 shows confounding by Z . Consider, for instance, a study of the effect of dreaming status (A) on growth hormone level (V), assuming that sleep stage (Z) is a cause of both. In that case, it would be invalid to derive dreaming status from sleep stage (DAG A). Nonetheless, the substitution would be valid if the confounding path could be blocked by conditioning on B (DAG B), which is an intermediary variable between Z and V . DAG C in Figure 4 shows confounding by a known, but unmeasured, U that would be blocked by conditioning on Z . Again, A_s should not be derived from Z in DAG C, but may be derived from Z in DAG D, because conditioning on B would remove the confounding bias. As an example, suppose that carotid atherosclerosis (Z) and coronary heart disease (A) affect vital status (V) and share a known, but unmeasured cause, say, tumor necrosis factor (U). In a study of the effect of coronary disease (A)

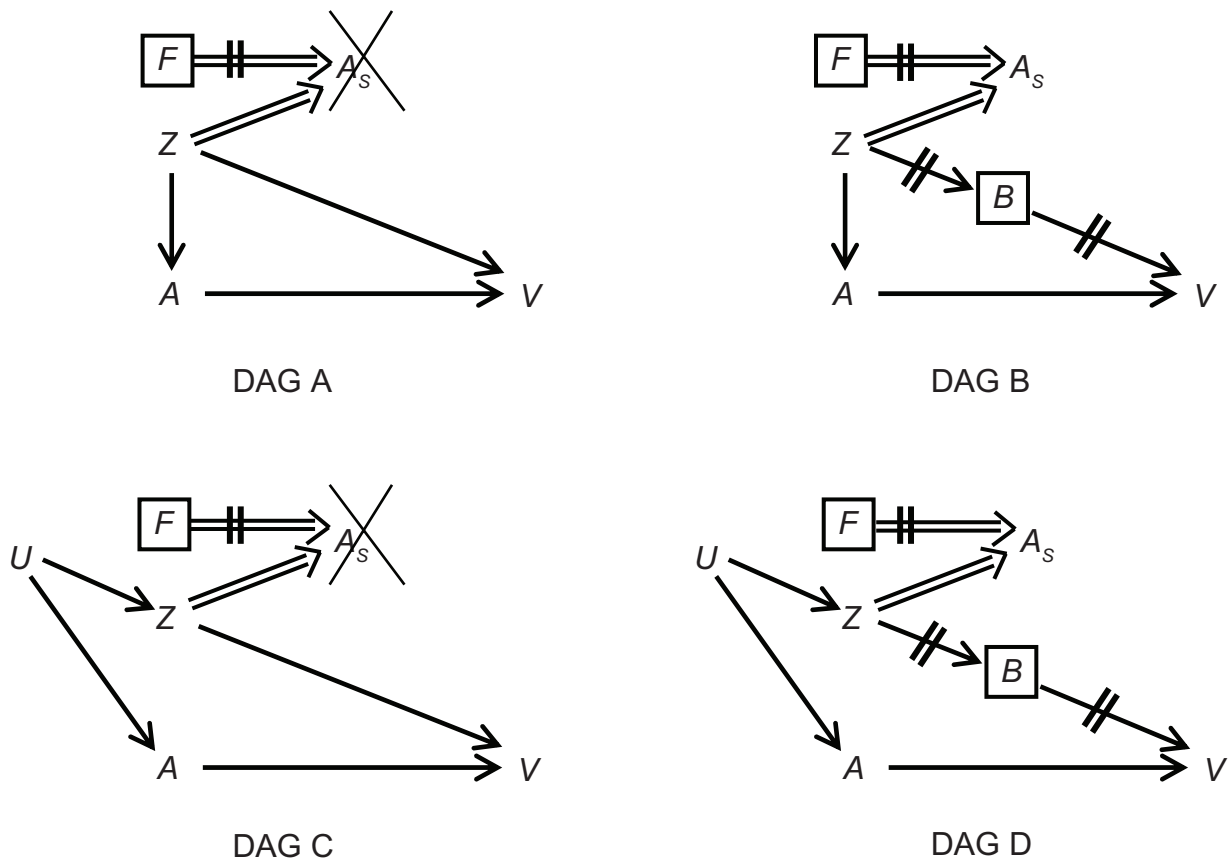


Figure 4 Examples of invalid and valid substitutions of A_s for A , when A_s is derived from Z .

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021 For personal use only.

on vital status (V), it would be invalid to derive the status of the coronary arteries from the level of carotid atherosclerosis (DAG C) unless all known directed paths from carotid atherosclerosis to vital status are blocked (DAG D).

Thought bias

Scientists impute the values of natural variables on a daily basis. But why? What property does a natural variable have that makes it scientifically interesting? A moment of reflection should provide the answer. A natural variable is an attribute of a physical object, whether that object is an electron, a cell, or a human being. As such, it is the cause of some other natural variables. The collection of natural variables makes up the ultimate DAG of the physical world (World 1), and knowing their effects is the essence of knowledge in many branches of science.

Natural variables come into existence via their causes, and are the reason for the existence (that is, they are the causes) of their effects. The bias we are about to describe is concerned with the fundamental question of existence: does a variable whose values we impute belong to the physical world? Might we erroneously think that some variable exists? If the answer to the last question is affirmative (ie, some variable does not exist), bias must be present in any study of that variable. Because the bias arose from an erroneous thought about existence, it may be called “thought bias” or “existence bias”.

Doubting existence is not a new idea in science. Physicists sometimes wonder whether objects, such as quarks, gravitons, or photons, indeed exist. Do the names describe real entities or are they merely inventions of the mind? Of course, if an object does not exist in physical reality, no attribute (variable) may be attached to it either. Contemporary biomedical researchers do not doubt that the objects of their research exist (cells, mice, people), but they sometimes question the existence of postulated attributes, such as disease states.^{16–20} Critics of a proposed medical entity have undoubtedly accepted the thesis that a variable might not exist. But even their opponents must have agreed with that thesis. Otherwise, they should not have felt the need to defend a particular syndrome.²¹

The claim that a variable exists in World 1 (or in World 2) should not be viewed as an independent theory. It is, however, a key element of many scientific (ie, causal) theories that involve that variable. When the claim of variable existence is false, theories in which the nonexistent variable plays a central role (eg, cause, effect, or effect modifier) are false as well.

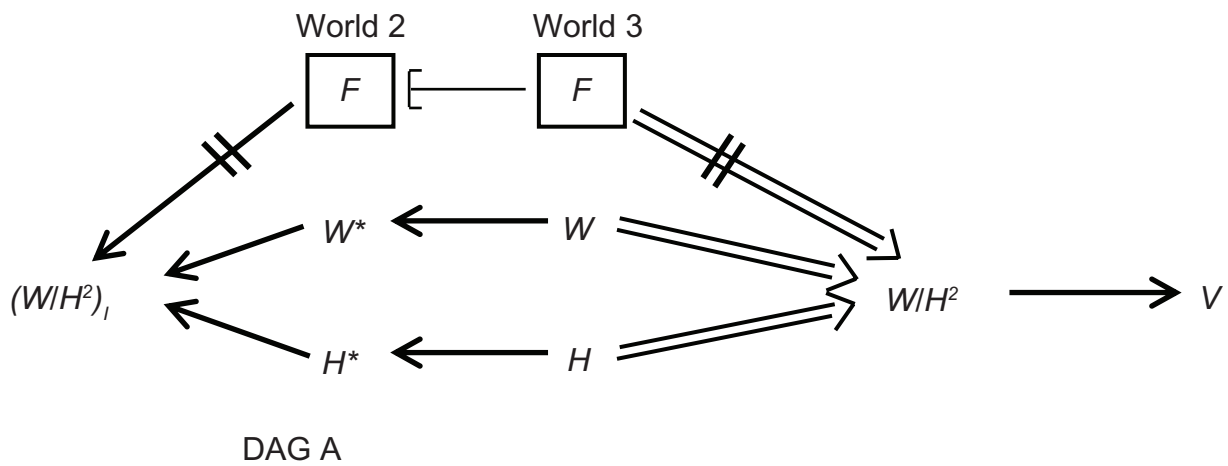
The falsehood attributed to thought bias is radically different from the falsehood due to other biases. We are not committing a mistake because the expected value of the estimator differs from the causal parameter (which may be null). Rather, we are committing a mistake because the causal parameter does not exist at all. The entire process of parameter estimation is a priori useless, because there is nothing to estimate. Stated figuratively, it is not the case of a causal arrow between two variables on which Nature might have written “The probability ratio is 1.0001” (or even “precisely 1”); it is the case of a nonexistent arrow. The distinction between null effect and no parameter might seem subtle, but it is crucial to the grasping of thought bias and its uniqueness.

Thought bias, type I: derived variables

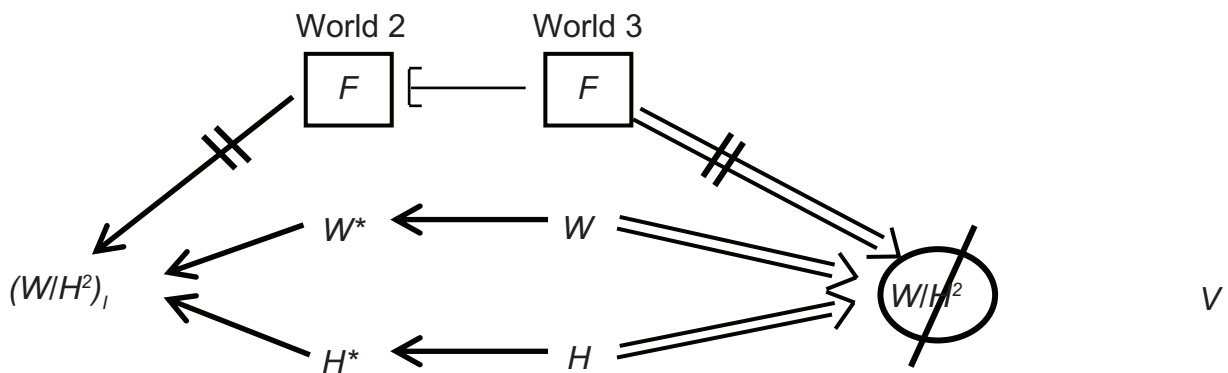
As we described earlier, scientists often derive variables to impute the values of natural variables, which are assumed to exist. Other times, however, they explicitly or implicitly assume that a derived variable could be the cause of World 1 variables. For example, biomedical researchers often impute the values of body mass index, $W/H^2 \ominus (W/H^2)$, on the assumption that true body mass index affects various outcomes^{22–25} (Figure 5, DAG A). That assumption is false, demonstrating a ubiquitous form of thought bias. The variable W/H^2 is not a natural variable, and therefore any study of its effects is affected by thought bias. The bias will be removed by recognizing that W/H^2 , a World 3 variable, does not exist in World 1 (Figure 5, DAG B). Again, the causal parameter here between W/H^2 and some outcome is not null, as might be the case for any two natural variables; it does not exist a priori. There is no causal parameter to estimate.

Note the subtle, but fundamental, distinction between derivation in the context of information bias and thought bias. In the former situation, no one claims that a theoretical derivation (D) is a natural variable. Rather, the derivation is coupled with a theory of why the unknown values of D substitute for the unknown values of a natural variable, A . For that reason D is renamed A_s . If A exists, information bias might be present and thought bias is absent. Here, in contrast, D itself, a mathematically made variable, is mistakenly assumed to be a natural variable that affects World 1 variables. In that case, thought bias is present and information bias is irrelevant, because the imputation of D itself ($D \ominus D_s$) generates useless information; the values of D are not attributes of an object in the physical world.

Our claim above is founded, of course, on the assumption that theoretical derivations are not natural variables, and they



DAG A



DAG B: Thought bias removed

Figure 5 Thought bias: type 1.

do not have effects in World 1. We will examine that thesis after the next section.

Thought bias, type 2: constructs

The second type of thought bias arises from claiming that a construct, a rough idea, is a natural variable of World 1. This type of thought bias is similar to the first type, but there are two differences, in that the values of constructs, unlike the values of derived variables, do not originate from a strict list of variables and are not derived by a specific function. Therefore, constructs have neither a uniform set of causes nor exact values. For instance, the construct called “athleticism” might be related to variables such as body fat, muscle volume, and speed, but the list could be modified (adding height, for instance), and there is no specific function by which its values might be derived.

An important distinction should be made between constructs (World 3) and ideas in the mind (World 2).

“Perceived athleticism”, for example, may be a World 2 variable, and thus, it may be a property of the perceiver. But “athleticism” is a World 3 variable; it does not describe a property of any person. “Perceived athleticism”, if it exists in World 2, may affect physical reality because perception (even self-perception) might influence behavior (eg, granting or denying privileges). “Athleticism”, however, is not a property of any person, and therefore, there are no causal parameters for its “effects”.

In this type of thought bias, the nonexistent variable is usually “imputed” from other variables, just as in type 1. In Figure 6, DAG A shows an example whereby two scientists have claimed to impute a construct they called “adiposity” (*A*) from weight (*W*) and height (*H*) by the function W/H^2 . Interestingly, that example of thought bias (according to our theory) was presented as an example of information bias.⁴ Although deciding between these two types of bias is not always simple, there are two clues.

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021 For personal use only.

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021
For personal use only.

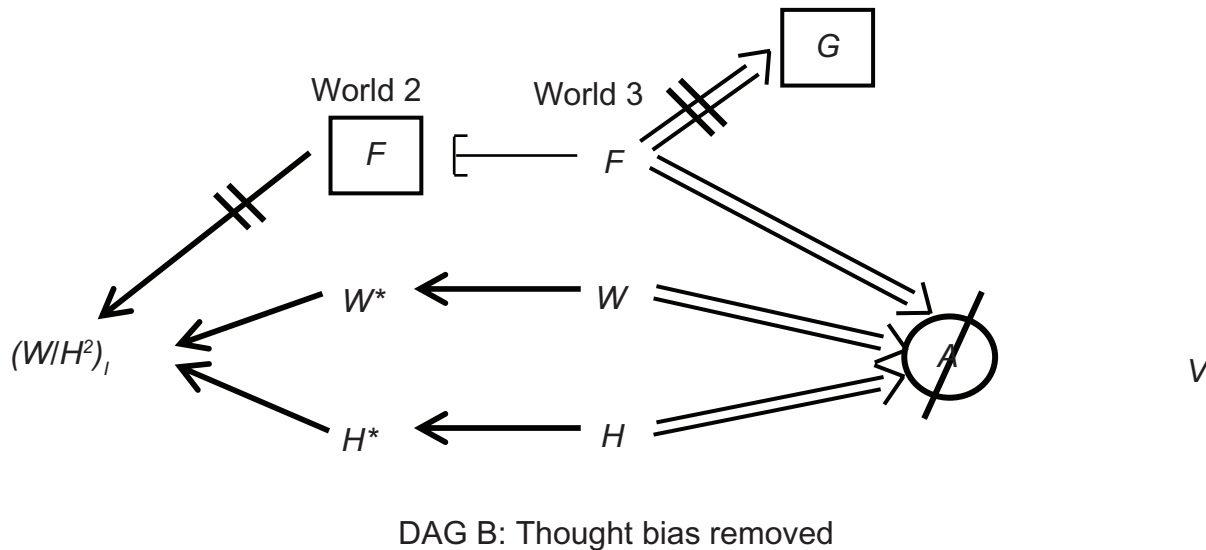
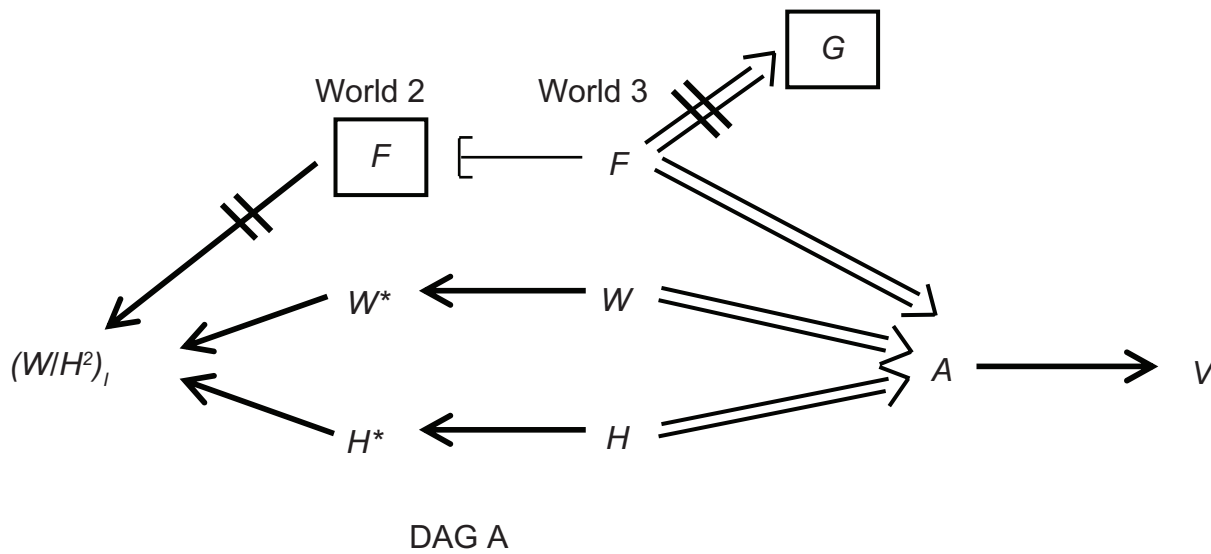


Figure 6 Thought bias: type 2. The arrow from F to G , with conditioning on G , indicates that the universe of all functions is often restricted to some subset.

First, the values of a construct are not grasped clearly and intersubjectively. Compare, for example, your grasping of something called “adiposity” (when it is not synonymous with “amount of fat tissue”) with your grasping of the variable “left ventricular volume”. Second, the “imputation” of a construct, unlike the imputation of a natural variable, is rarely founded on an imputation theory. For instance, why should we accept the claim that “adiposity” is imputed by W/H^2 rather than the claim that it is imputed by some other function of these, or other, variables?⁵ In fact, the scientists who presented that example wrote that “ $\text{kg}/\text{m}^{2.5}$ [$W/H^{2.5}$] might be a better choice” for capturing the values of A .⁴ To which we respond: might “ $e^{\sqrt{W/H}}/H^{3.7296\pi}$ ” be a better choice?

Do theoretical derivations cause anything in World I?

To substantiate the claim that a theoretical derivation is not the cause of any natural variable, we consider the simple case where D is derived from two natural variables. The arguments hold for derivations from any number of variables, including just one.

Figure 7 shows three possible causal structures that involve three natural variables, X , Y , and V ; a derived variable, $D^{x,y}$; and the universe of all functions, denoted F . Note a key point here, which might be misunderstood by many: F is not a means to estimate the effect of X and Y on V ; it is another cause of V besides X and Y (via the path $F \Rightarrow D^{x,y} \rightarrow V$).

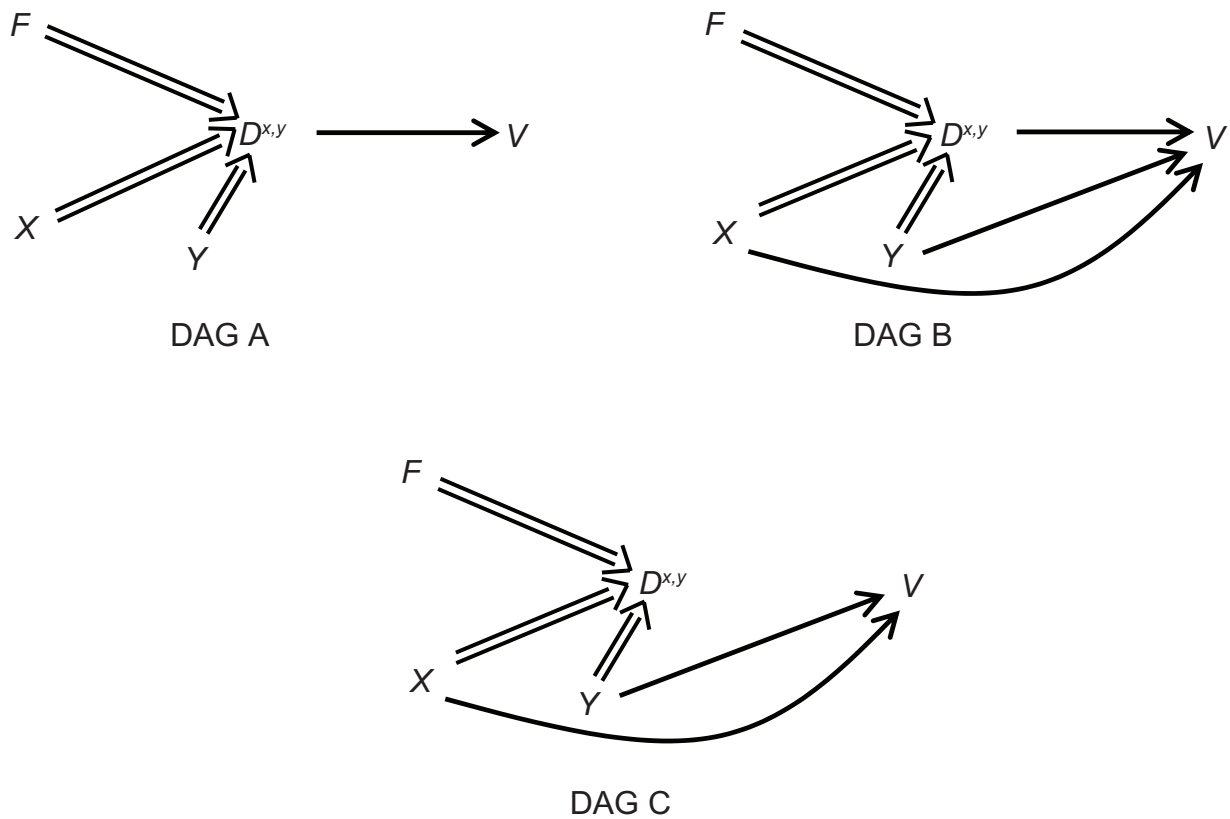


Figure 7 Alternative structures for the causal relation of V (a natural outcome variable) with X (natural variable), Y (natural variable), F (universe of all functions of X and Y), and $D^{x,y}$ (a variable derived from X and Y).

According to DAG A (Figure 7), $D^{x,y}$ is a cause of V , and a mediator of the effects of three variables on V , namely, F , X , and Y . If this structure is assumed to describe causal reality, we should accept several troubling consequences:

First, neither X nor Y has any direct effect on V . These two natural variables affect a third natural variable, V , only via another variable, which is a mathematical derivation from their values. They do not affect V by any other causal path.

Second, the derived variable, $D^{x,y}$, is an indispensable link for the association between pairs of natural variables. Without it, neither X nor Y would have been associated with V . Because $D^{x,y}$ is also an effect of F , we observe bizarre situations, such as the following. Let F take the value “ $X + 0 \cdot Y$ ” (ie, for any value of X and Y , the function tells us that the derived variable, $D^{x,y}$, should simply be X). Then, X is associated with V solely because $D^{x,y} (= X)$ is associated with V , which is circular reasoning: X is associated with V because $X (= D^{x,y})$ is associated with V . The oddity of the inference originates from the causal assumption of DAG A, namely, there is no natural path by which X affects V .

Third, the causal path $F \Rightarrow D^{x,y} \rightarrow V$ implies that the universe of functions affects a natural variable. We should conclude, for instance, that someone’s chances of dying

would be different depending on the existence of a particular function.⁷ This inference is, of course, absurd. We do not assume that an idea by itself (a World 3 object) can affect vital status, unless it is incorporated into some mind (World 2).

Finally, the causal path $F \Rightarrow D^{x,y} \rightarrow V$ is not testable. Given the nature of the variable F , whose values are functions, it is impossible to study the “association” between V and F . Every person takes a single value of V at every moment, but all of the values of F coexist.

An alternative causal structure is shown in DAG B (Figure 7). As in DAG A, the derived variable, $D^{x,y}$, is still assumed to be a cause of a natural variable, V , but its makers, the natural variables X and Y , are assumed to affect V via paths that do not pass through $D^{x,y}$. In this diagram, the association of $D^{x,y}$ with V contains confounding by X and Y , which should be removed in order to estimate the presumed effect $D^{x,y} \rightarrow V$. But deconfounding by conditioning on X and Y is impossible. Conditioning would fix the value of $D^{x,y}$, preventing $D^{x,y}$ from taking different values, which is essential for effect estimation. We could try to circumvent the problem by conditioning on other variables along the confounding paths, instead of conditioning on X and Y (eg, on L in $X \rightarrow L \rightarrow V$). But in this case, DAG B is simplified to DAG

A after deconfounding, and shares some of the problems of DAG A. DAG B leads to another disturbing conclusion: it is impossible to distinguish between the effects of X and Y on V via $D^{x,y}$ and their “direct” effects (via other pathways). We have shown the impossibility of estimating these direct effects when D was a change variable (ΔX).⁷ Other modifications of DAG A or DAG B, which retain the arrow $D^{x,y} \rightarrow V$, are subject to similar problems.

In summary, we propose to reject DAG A and DAG B and accept DAG C as the true causal structure of World 1: $D^{x,y}$ does not cause anything in the physical world. Neither $D^{x,y}$ nor the arrow $D^{x,y} \rightarrow V$ exists. There is nothing to estimate.

Three pairs of bias

Although not widely appreciated, confounding bias²⁶ and colliding bias²⁷ may be viewed as a pair of antithetical biases. Table 1 shows numerous examples of their opposite characteristics. For example, many characteristics of colliding bias are no more than rewriting the characteristics of confounding bias, using an opposite term: “effect” instead of “cause”, “several” instead of “single”, “closed” instead of “open”, “conditioning” instead of “not conditioning”, and so on.

Similarly, thought bias may be viewed as antithetical to information bias (Table 2). The main feature of information bias may be described as having the wrong values of a variable, whereas the main feature of thought bias may

be described as having the wrong variable. A variable that does not exist is wrong, because it is not linked to any causal parameter; it is not part of the causal structure of the physical world.

Information bias is removed whenever A_I provides perfect information on A , whereas thought bias is removed whenever the so-called information on A is ignored. It is possible to depict the presence of information bias in a DAG, but not its removal. Conversely, it is possible to depict the removal of thought bias, but not its presence. Both types of bias are founded on an untestable assumption (an axiom). The axiom of information bias states that the arrow depicting imprinting exists ($A \rightarrow A^*$), whereas the axiom of thought bias states that no arrow emanates from a nonexistent A .

Of note, information bias follows attempts to remove other types of bias, whereas thought bias is antecedent to other types of bias. A variable that does not exist cannot create, and cannot remove, confounding bias, colliding bias, or information bias. Likewise, it cannot create or remove effect-modification bias²⁸ and causal-pathway bias,⁹ a third pair of antithetical biases. Of the six types of bias, four are concerned with causal arrows, ie, confounding, colliding, effect-modification, and causal pathway. Thought bias and information bias are concerned, respectively, with variables and values, which are the foundations of every study.

Table 1 Antithetical characteristics of confounding bias and colliding bias

	Confounding bias	Colliding bias
Main feature	Common (shared) cause of two variables	Common (shared) effect of two variables
Natural causal structures	Single: $E \leftarrow C \rightarrow D$	Several. For example: $E \rightarrow C \leftarrow D$ $E \rightarrow C \leftarrow L \rightarrow D$
Causal path before conditioning	Associational (open)	Blocked (closed)
Presence of bias	Not conditioning on one of the confounders (of the relevant association)	Conditioning on all of the colliders (on the relevant path)
Magnitude of bias	Depends on: 1. the magnitude of C 's effects on E and D 2. the distribution of C	Depends on: 1. the magnitude of effect modification between the causes of C 2. the value of C
Removal of bias (primary method)	Conditioning on all of the confounders (of the relevant association)	Not conditioning on at least one collider (on each relevant path)
Removal of bias (secondary method)	Conditioning on an effect of a confounder (on the relevant path), rather than conditioning on the confounder. For example: $E \leftarrow C \rightarrow \square \rightarrow D$	Conditioning on a cause of a collider (on the relevant path), in addition to conditioning on the collider. For example: $E \rightarrow \square \leftarrow \square \rightarrow D$

Notes: The effect of interest is $E \rightarrow D$. C denotes a confounder or a collider.

Pragmatic and Observational Research downloaded from https://www.dovepress.com/ by 35.172.136.29 on 31-Jul-2021 For personal use only.

Table 2 Antithetical characteristics of information bias and thought bias

	Information bias	Thought bias
Main feature	A_i has the wrong values	A is the wrong variable
Reason for bias	A exists in World 1 (or in World 2), but its values are unknown	A does not exist in World 1 (nor in World 2); it has no values
Presence of bias	Using A_i when its values differ from A 's values	Using A_i when A does not exist
Removal of bias	A_i actually provide perfect information on A	A_i is assumed to provide worthless (ignored) information on A
Causal path	Surrogate path for the causal path of interest	Nonexistent "causal path of interest"
Causal diagram	Can depict its presence, but not its removal	Can depict its removal, but not its presence
Untestable assumption (axiom)	$A \rightarrow A^*$ (imprinting)	$\textcircled{A} \vee$
Relation to other biases	Succedent	Antecedent

To our knowledge, that organized structure of six biases in three antithetical pairs has not been described. Moreover, textbooks of epidemiology, biostatistics, and medical research typically highlight only three categories of bias, ie, confounding, colliding bias (under various names), and information bias (under various names).

Discussion

As this article shows, causal diagrams proved to be an essential methodological tool in scientific inquiry. Not only did they offer deeper insight into the ideas of "measurement" and "information", but they also explained how (and when) the unknown true values of a natural variable may be imputed from information about other natural variables.

Causal diagrams,²⁶ along with Popper's description of World 1 (the physical world), World 2 (the mind), and World 3 (the autonomous world of ideas),⁸ have also helped us to identify a previously unrecognized bias, which may be called thought bias or existence bias. Before asking about the effect of one variable on another, we should pause for a moment and ask whether both variables exist in World 1 or World 2. If the answer is "no", that pair is not of interest in science. No causal parameter is linked to a nonexistent variable.

Of the two types of thought bias, the first type has severely affected all branches of science. Countless estimated effects of derived variables, which show up regularly on the pages of scientific journals, are false a priori; the causal parameters do not exist. Again, we refer here to nonexistent causal parameters of theoretical derivations (D). An empirical derivation (D_i) is a natural variable by itself, which can affect other natural variables, regardless of how well D_i resembles the unknown values of D . For example, computed body mass index (D_i) may affect one's behavior and its effects, even though true body mass index (D) is not linked to any causal parameter in the world of natural variables.

Many scientists have not recognized the crucial distinction between natural variables (A); theoretical derivations that substitute for natural variables (A_s); and useless theoretical derivations (D). The third category is common: weight divided by height squared; definitions of hypertension; metabolic syndrome status; cognitive function at one time minus cognitive function at another time; and number of apneas and hypopneas divided by hours of sleep, to name a few examples from biomedical science. Examples from physics might include variables such as acceleration, torque, and potential energy. In retrospect, an unimaginable number of articles in all branches of science are utterly useless; they are no more than a mathematical exercise on derived variables, disconnected from causal reality.

Thought bias did not spare even methodological articles on causal inquiry. Three authors claimed that "cause-specific mortality variables" are causes of death, and displayed two nonexistent causal arrows: death from tumor \rightarrow death; death from heart attack \rightarrow death.²⁷ These variables are derived from vital status and underlying disease status; they do not exist in World 1 as causes of death. (Tumor status may affect vital status, but death from tumor may not.) Other authors displayed derivations from natural variables, labeled "sufficient causes", as causes of other natural variables.^{29,30} Sufficient causes (a false idea under indeterminism⁸) belong to World 3; they are derived from natural variables and do not exist in the physical world as proximal causes of natural effects.

The entrenched mistake may be traced to several sources. First, in statistics, there is no distinction between natural variables and derived variables. X/Y is a legitimate statistical variable, just like X and Y . Second, functions are regularly used to estimate effects, and it is easy to miss the subtle distinction between the using of a function to estimate an effect and claiming that the output of functions has an effect. Formal

decoding of causal assertions by DAGs, as shown here, has revealed the absurdity of the latter claim. Third, empirical derivations of categoric variables (categoric-type D_i) are essential for human actions. A practicing physician decides whether to prescribe an antihypertensive drug based on a dichotomy of measured blood pressure (ie, based on classified hypertension status). But the need to derive a variable for medical practice does not imply that its theoretical counterpart must have a place in causal inquiry. If hypertension status (D), as derived from true blood pressure, does not substitute for the values of some natural variable (A), there is no merit in estimating its nonexistent causal parameters, and there is no merit in studying how other variables affect it. Again, we emphasize the fundamental distinction between nonexistent causal parameters of the derived variable whose values are unknown (D), and the (unrelated) causal parameters of its empirical version (D_i). Actual classification of a patient as hypertensive or normotensive (D_i being classified hypertension status) may have numerous effects in World 1.

Before recognizing the uniqueness of thought bias, we mistakenly considered it to be another example of confounding bias.⁵⁻⁷ One might claim, for example, that finding a null association between a derived variable and some presumed effect, after the necessary conditioning, is simply evidence for confounding bias: the causal parameter is null, rather than nonexistent. We offer a different viewpoint. The basic causal structure of World 1 contains time-indexed natural variables, all of which are connected with causal arrows pointing in the direction of the time axis. Some of the arrows carry large causal parameters; some small (or even nearly null) parameters; and some null parameters. In principle, a null parameter could have been non-null and vice versa, but the structure is otherwise fixed. A theory of confounding bias may be invoked whenever we attempt to estimate a causal parameter within this structure. But derived variables do not show up at all in that diagram. They are never causes of natural variables, and therefore, the concept of confounding bias does not apply. Thought bias is antecedent to confounding bias and to any other bias as well.

The absence of derived variables (and their “causal parameters”) from the DAG of natural variables may be considered a rule of causation, which must be accepted axiomatically. We supplied arguments in favor of this axiom, but we cannot deduce it. Alternatively, one may assume that all derived variables have a null effect on natural variables, regardless of the evidence provided against it. In contrast, our axiom allows us to dismiss the empirical evidence altogether and the absurd inference that follows. Therefore, any study

that estimates the effect of a derived variable (including the null) is ridden with bias.

Rejection of our DAG-based thesis would require both a claim and a competing proposal: a claim that thought bias does not exist (in any possible study), and a proposal for the demarcation between a legitimate derivation and a useless derivation. There is no middle ground. Otherwise, any conceivable derivation of one variable from a set of other variables might be given a name and claimed to be a cause of some effect. For example, one may claim that $RINDEX = e^{\sqrt{WEIGHT}^{3.274} + 0.5 BLOOD_PRESSURE^{2.1} \ln(GLUKCOSE)} / \log_x CIGARETTES$ is a cause of stroke and estimate the effect of $RINDEX$ on stroke.

It is not enough to respond by saying that “the derivation should make sense”. What exactly is the “sense” in that assertion? How do you propose to formally distinguish between sense and senseless? We argued here that a theoretical derivation makes sense only if it serves as the first step in the imputation of the values of a natural variable. Any theoretical derivation that fails to meet this criterion should be rejected a priori as thought bias. No causal knowledge may be gained by imputing its values.

Disclosure

The authors report no conflicts of interest in this work.

References

- Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995; 82(4):669–688.
- Hernán MA, Robins JM. A structural approach to observation bias. *Am J Epidemiol*. 2005;161 Suppl:S100.
- Shahar E. Causal diagrams for encoding and evaluation of information bias. *J Eval Clin Pract*. 2009;15(3):436–440.
- Hernán MA, Cole SR. Causal diagrams and measurement bias. *Am J Epidemiol*. 2009;170(8):959–962.
- Shahar E. The association of body mass index with health outcomes: Causal, inconsistent, or confounded? *Am J Epidemiol*. 2009;170(8): 957–958 (rejoinder 963–964).
- Shahar E. Metabolic syndrome? A critical look from the viewpoints of causal diagrams and statistics. *J Cardiovasc Med (Hagerstown)*. 2010; 11(10):772–779.
- Shahar E, Shahar DJ. Causal diagrams and change variables. *J Eval Clin Pract*. Sep 12 2010. [Epub ahead of print].
- Popper KR. *The Open Universe: An Argument for Indeterminism*. London, UK: Routledge; 1988.
- Shahar E, Shahar DJ. On the causal structure of information bias and confounding bias in randomized trials. *J Eval Clin Pract*. 2009;15(6): 1214–1216.
- Atchley WR, Anderson D. Ratios and the statistical analysis of biological data. *Syst Zool*. 1978;27(1):71–78.
- Kronmal RA. Spurious correlation and the fallacy of the ratio standard revisited. *J R Stat Soc, Series A*. 1993;156(3):379–392.
- Spiertes P. Variable definition and causal inference. In: Glymour C, Wang W, Westerstahl D, editors. Proceedings of the 13th International Congress of Logic Methodology and Philosophy of Science. London, UK: College Publications, King's College; 2009.

13. Gale EA. The myth of the metabolic syndrome. *Diabetologia*. 2005; 48(9):1679–1683.
14. Kucharska-Newton AM, Couper DJ, Pankow JS, et al. Diabetes and the risk of sudden cardiac death, the Atherosclerosis Risk in Communities study. *Acta Diabetol*. Oct 24 2009. [Epub ahead of print].
15. Popper KR. *The Open Society and its Enemies* (Volume II). Princeton, NJ: Princeton University Press; 1971.
16. Serratrice G. Does fibromyalgia exist? *Rev Rhum Mal Osteoartic*. 1990;57(3):260–266. French.
17. Bohr TW. Fibromyalgia syndrome and myofascial pain syndrome. Do they exist? *Neurol Clin*. 1995;13(2):365–384.
18. Wallace PG. Post-viral fatigue syndrome. Epidemiology: A critical review. *Br Med Bull*. 1991;47(4):942–951.
19. Ling H, O'Sullivan SS, Holton JL, et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. *Brain*. 2010;133(Pt 7): 2045–2057.
20. Aringer M, Kuhn A. Does cutaneous lupus erythematoses really exist? *Z Rheumatol*. 2010;69(5):389–392. [German].
21. Grundy SM. Does the metabolic syndrome exist? *Diabetes Care*. 2006; 29(7):1689–1692.
22. Bhattacharya S, Campbell DM, Liston WA, Bhattacharya S. Effect of body mass index on pregnancy outcomes in nulliparous women delivering singleton babies. *BMC Public Health*. 2007;7:168.
23. Russo MJ, Hong KN, Davies RR, et al. The effect of body mass index on survival following heart transplantation: Do outcomes support consensus guidelines? *Ann Surg*. 2010;251(1):144–152.
24. Sestak I, Distler W, Forbes JF, Dowsett M, Howell A, Cuzick J. Effect of body mass index on recurrences in tamoxifen and anastrozole treated women: An exploratory analysis from the ATAC trial. *J Clin Oncol*. 2010;28(21):3411–3415.
25. Hart CL, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: Analysis of data from two prospective cohort studies. *BMJ*. 2010;340:c1240.
26. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, UK: Cambridge University Press; 2009.
27. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615–625.
28. Shahar E. Estimating causal parameters without target populations. *J Eval Clin Pract*. 2007;13(5):814–816.
29. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol*. 2007;166(9):1096–1104.
30. Gatto NM, Campbell UB. Redundant causation from a sufficient cause perspective. *Epidemiol Perspect Innov*. 2010;7:5.

Pragmatic and Observational Research

Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>

system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress