Causal Directed Acyclic Graphs and the Direction of Unmeasured Confounding Bias

Tyler J. VanderWeele,[a] *Miguel A. Hernán,*[b] and *James M. Robins*[b,c]

Abstract: We present results that allow the researcher in certain cases to determine the direction of the bias that arises when control for confounding is inadequate. The results are given within the context of the directed acyclic graph causal framework and are stated in terms of signed edges. Rigorous definitions for signed edges are provided. We describe cases in which intuition concerning signed edges fails and we characterize the directed acyclic graphs that researchers can use to draw conclusions about the sign of the bias of unmeasured confounding. If there is only one unmeasured confounding variable on the graph, then nonincreasing or nondecreasing average causal effects suffice to draw conclusions about the direction of the bias. When there are more than one unmeasured confounding variable, nonincreasing and nondecreasing average causal effects can be used to draw conclusions only if the various unmeasured confounding variables are independent of one another conditional on the measured covariates. When this conditional independence property does not hold, stronger notions of monotonicity are needed to draw conclusions about the direction of the bias.

(*Epidemiology* 2008;19: 720–728)

Control for confounding variables is one of the central challenges of epidemiologic studies. Directed acyclic graphs that represent causal relations among variables have been used extensively to determine the variables on which it is necessary to condition to control for confounding in the estimation of causal effects.[1–4] However, control for confounding is often inadequate when certain variables that are known to be confounders are not measured in a particular study. In such cases it is sometimes possible to provide bounds on the magnitudes of the true causal effects.[5–7] Alternatively, certain sensitivity analysis techniques can sometimes be used to assess the impact of the unmeasured confounding variables.[8–13] Some of these sensitivity techniques are model-dependent. In this paper we present results that allow a researcher in certain circumstances to determine the sign of the bias arising when control for confounding is inadequate. The results are not model-dependent and can be used to draw conclusions about the presence of a true causal effect without the use of sensitivity analysis. Sensitivity analysis may, however, still be useful in such cases for drawing conclusions about likely upper bounds for the magnitude of the effect. In Appendix 1, we present some related theory that was developed elsewhere.[14] That work required fairly strong monotonicity assumptions, which are discussed further below. In this paper we employ weaker monotonicity assumptions—only the presence of nonincreasing or nondecreasing *average* causal effects—and characterize those graphs on which these weaker assumptions are sufficient to determine the sign of the bias of unmeasured confounding.

Consider a study examining the effect of a potentially beneficial exposure A on some disease Y. Suppose also that an individual's likelihood of exposure depends on the individual's state of health U, and that the individual's state of health also affects the likelihood of developing the disease. The directed acyclic graph corresponding to these causal relationships is given in Figure 1 (this example uses concepts from causal inference that will be made more precise in the following section). A valid estimate of the causal effect of the exposure can be computed by controlling for U. Suppose now that data on the state of health of the study subjects is not available. Without controlling for U, the relationship between the exposure and disease is confounded and the observed risk difference does not equal the causal risk difference. Under the assumptions given in this paper it is possible to rigorously show that if less healthy individuals have a higher probability of receiving the exposure and if less healthy individuals are also more likely to develop the disease, then the estimate of the risk difference not controlling for U is in fact conservative for the true causal effect of A on Y.

This result is unsurprising; it is what we would expect intuitively. However, we will present other examples for which intuition breaks down. It is therefore important to have a rigorous theory describing when conclusions about the sign of the bias of unmeasured confounding can be drawn. The results are given within the context of causal directed acyclic graphs. We will introduce some new concepts, including

Submitted 20 July 2007; accepted 24 January 2008; posted 14 July 2008.
From the [a]Department of Health Studies, University of Chicago, Chicago, IL; and Departments of [b]Epidemiology and [c]Biostatistics, Harvard School of Public Health, Boston, MA.
Supplemental material for this article is available with the online version of the journal at www.epidem.com; click on "Article Plus."
Correspondence: Tyler J. VanderWeele, Department of Health Studies, University of Chicago, 5841 S. Maryland Ave., MC 2007, Chicago, IL 60637. E-mail: vanderweele@uchicago.edu.

ISSN: 1044-3983/08/1905-0720
DOI: 10.1097/EDE.0b013e3181810e29

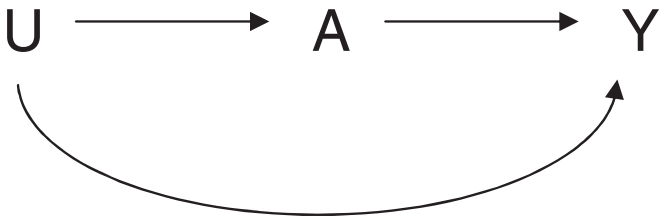

FIGURE 1. Example illustrating confounding by health status. Y indicates disease; A, exposure; U, unmeasured health status.

minimal causal directed acyclic graphs, signed causal directed acyclic graphs, and various notions of monotonic effects.

Causal Directed Acyclic Graphs

We begin by reviewing definitions and some central results concerning causal directed acyclic graphs. A *directed acyclic graph* is composed of variables (nodes) and arrows between nodes (directed edges) such that the graph is acyclic—ie, such that it is not possible to start at any node, follow the directed edges in the arrowhead direction and end up back at the same node. A *causal* directed acyclic graph is one in which the arrows can be interpreted as causal relationships and in which all common causes of any pair of variables on the graph are also included on the graph. If there is a directed edge from A to Y then A is said to be a parent of Y and Y is said to be a child of A. Additional details concerning causal directed acyclic graphs can be found in the work of Greenland et al.[2] Greater formalization is provided by Pearl[1,15] and Spirtes et al.[16] By representing causal relations, causal directed acyclic graphs encode the causal determinants of statistical associations.

Statistical associations on causal directed acyclic graphs can arise in a number of ways. Two variables, A and B, may be statistically associated if A is a cause of B or if B is a cause of A. Even if neither is the cause of the other, the variables A and B may still be statistically associated if they have some common cause C. Finally, the variables A and B may be statistically associated if they have a common effect K and the association is computed within strata of K.

More formally, the statistical association between variables can be determined by blocked and unblocked paths. A path is a sequence of nodes connected by edges regardless of arrowhead direction; a directed path is a path that follows the edges in the direction indicated by the graph's arrows. If there is a directed path from A to Y then A is said to be an ancestor of Y and Y is said to be a descendent of A. A collider is a particular node on a path such that both the preceding and subsequent nodes on the path have directed edges going into that node, ie, both the edge to and the edge from that node have arrowheads into the node. A path between A and B is said to be blocked given some set of variables Z if either there is a variable in Z on the path that is not a collider or if there is a collider on the path such that neither the collider itself nor any of its descendants are in Z. If all paths between A and B are blocked given Z, then A and B are said to be d-separated given Z. It has been shown that if A and B are d-separated given Z, then A and B are conditionally independent given Z.

One further result regarding directed acyclic graphs has proved particularly useful in determining whether a particular set of variables (or none at all) suffice to control for confounding when estimating the causal effect of some exposure A on some outcome Y. Let Y_a denote the counterfactual variable Y intervening to set the exposure variable A, possibly contrary to fact, to level a. The causal effect of A on Y comparing 2 levels of A, a_0 and a_1 say, is defined simply as the causal risk difference $\mathbb{E}[Y_{a_1}] - \mathbb{E}[Y_{a_0}]$. Following Pearl,[15] we will refer to $\mathbb{E}[Y_a]$ as the causal effect of intervening to set A to a. The backdoor path adjustment theorem[1] states that for intervention variable A and outcome Y, if a set of variables Z such that no variable in Z is a descendent of A blocks all "back-door paths" from A to Y (ie, all paths with directed edges into A) then conditioning on Z suffices to control for confounding for the estimation of the causal effect of A on Y and this causal effect is given by:

$$\mathbb{E}[Y_a] = \sum_z \mathbb{E}[Y|A = a, Z = z]P(Z = z).$$

Note that the expression on the right hand side of the equation is simply a standardized mean. Pearl's backdoor path adjustment result is a graphical generalization of Theorem 4 of Rosenbaum and Rubin[17] and a special case of Robin's g-formula.[18,19]

Minimal Causal Directed Acyclic Graphs

We need one further definition. Consider some exposure A and some outcome Y. Let X be some set of nondescendents of A for which we might control. We will say that a causal directed acyclic graph is minimal with respect to A, Y, and X if the variables on the causal directed acyclic graph consist only of A, Y, X, and all variables that are common causes of any 2 variables in $\{A, Y, X\}$. Recall that the requirement for a directed acyclic graph to be causal is that every common cause of any 2 variables on the graph must also be on the graph. Thus a causal directed acyclic graph that is minimal with respect to A, Y, and X is the smallest causal directed acyclic graph that contains A, Y, and X. Furthermore, any causal directed acyclic graph with A, Y, and X can be reduced to one that is minimal with respect to A, Y, and X by eliminating all variables other than A, Y, X, and their common causes. As an example, consider causal directed acyclic graph given in Figure 2. This is not minimal with respect to A, Y, and X but can be made so by eliminating the variables Z and Q from the graph. Note that neither Z nor Q is a common cause of any 2 other variables on the graph. The resulting graph is shown in Figure 3. The graph in Figure 3 is minimal with respect to A, Y, and X. Thus a detailed causal directed

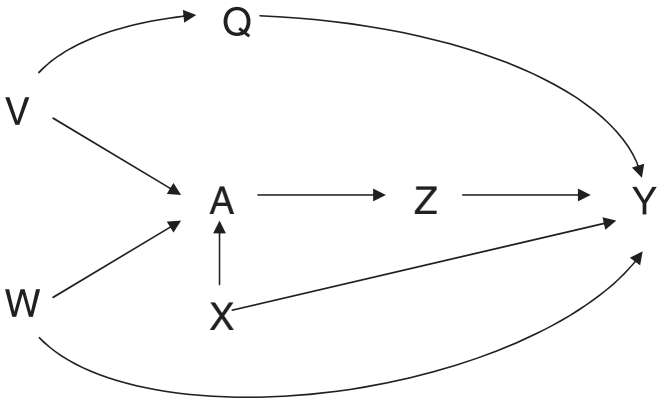

FIGURE 2. Graph that is not minimal with respect to A, Y, and X. Y indicates outcome; A, indicates exposure; X, indicates measured confounding variable; V and W, indicate unmeasured confounding variables; Q and Z, indicate intermediate variables.

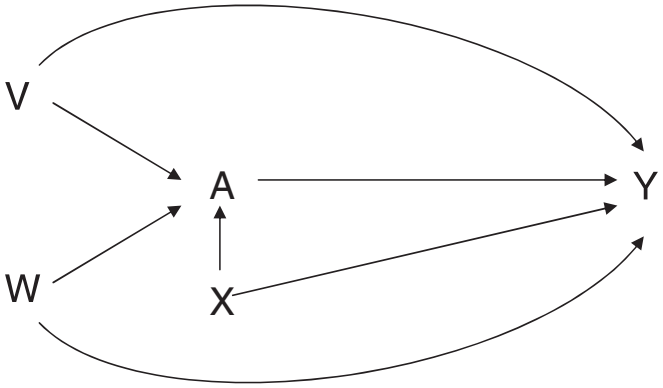

FIGURE 3. Graph that is minimal with respect to A, Y, and X. Y indicates outcome; A, indicates exposure; X, indicates measured confounding variable; V and W, indicate unmeasured confounding variables.

acyclic graph might involve several intermediate variables that would not be on a graph that is minimal. A graph that is minimal with respect to A, Y, and X will have considerably less detail. In particular, a graph that is minimal with respect to A, Y, and X has the following features: (i) there are no intermediate variables on the graph between A and Y; (ii) for every common cause C_i of A and Y there are no intermediate variables between C_i and A other than X or possibly other common causes of A, Y, and X; and (iii) for every common cause C_i of A and Y there are no intermediate variables between C_i and Y other than A, X and possibly other common causes of A, Y and X.

Signed Causal Directed Acyclic Graphs

Our results concerning the sign of the bias of unmeasured confounding will be stated in terms of signed edges. When signs are given to edges of a directed acyclic graph various counterintuitive results can sometimes arise. It is thus important to define precisely what we mean by a signed edge and to understand what conclusions we can draw from signed edges.

Signs might be given to edges to indicate a variety of relationships. For example, a sign might be given to an edge to indicate that intervening on the parent will increase or leave unchanged the average value of the child over the population. This is a nondecreasing average causal effect; in this setting a particular intervention to increase one variable would thus either increase or leave unchanged the average value of the outcome over the whole population. This is a relatively weak condition for giving a sign to an edge, and it is this weak condition that we will consider in this and the next section. We will consider stronger conditions in a subsequent section and in Appendix 1. In general, whether a sign can be appropriately placed on an edge depends also on the context ie, on which other variables are present on a directed acyclic graph. Consider some variable S that is a parent of some other variable T, and let Q denote the parents of T other than S. We will say that S has a positive average monotonic effect on T if increasing S with Q fixed always increases or leaves unchanged the average value of T over the population. More formally, S has a positive average monotonic effect on T if $\mathbb{E}[T|S, Q]$ is nondecreasing in S for all values of Q. Similarly, S has a negative average monotonic effect on T if $\mathbb{E}[T|S, Q]$ is nonincreasing in S for all values of Q. Whether S has a positive average monotonic effect on T depends on which parents of T are on the graph. The variable S might have a positive average monotonic effect on T on one directed acyclic graph but not on another graph that has more parents of T.

When a parent S has a positive average monotonic effect on child T, we will say that the $S - T$ edge is of positive sign. When S has a negative average monotonic effect on T, we will say that the edge is of negative sign. If S has neither a positive average monotonic effect nor a negative average monotonic effect on T, then the edge is said to be without sign. The sign of a path is then defined to be the product of the signs of the edges that constitute that path. If one of the edges on a path is without a sign then the sign of the path is said to be undefined. A signed causal directed acyclic graph is a causal directed acyclic graph with signs on those edges that allow them.

Sign of the Bias of Unmeasured Confounding

We can now consider the sign of the bias that arises when control for confounding is inadequate. As noted above, Pearl[1] showed that for intervention variable A and outcome Y, if a set of variables Z such that no variable in Z is a descendent of A blocks all back-door paths from A to Y, then the expected value of Y intervening to set $A = a$ is given by

$$\mathbb{E}[Y_a] = \sum_z \mathbb{E}[Y|A = a, Z = z]P(Z = z). \quad (1)$$

Now if X is some set of variables that does not block all backdoor paths from A to Y and an attempt is made by using the analog of the correct formula to estimate the causal effect on Y of intervening to set $A = a$ controlling only for X, one would obtain

$$S_a = \sum_x \mathbb{E}[Y|A = a, X = x]P(X = x). \quad (2)$$

Expression (1) will in general differ from expression (2) since expression (2) does not control for all the confounding variables. Result 1 relates signed edges to the sign of the bias that arises when control for confounding is inadequate. The proof of Result 1 is given in Appendix 2; the proof makes use of a result concerning potential outcomes.[20] The result requires that the intervention variable A be binary (an assumption which we discuss further below).

Result 1. Suppose that a directed acyclic graph is minimal with respect to A, Y, and X, where A is binary and X is a set of nondescendents of A. Let U denote the nondescendents of A on the graph other than X. Suppose further that if U contains more than one variable, $U = (U_1, \ldots, U_n)$, then the components of U are conditionally independent given X. The following statements then hold:

i. If for each U_i, the $U_i - A$ edge (if it exists) is of the same sign as the $U_i - Y$ edge (if it exists) then $S_1 \geq \mathbb{E}[Y_1]$ and $S_0 \leq \mathbb{E}[Y_0]$. That is, the estimate of the causal effect on Y of intervening to set $A = 1$ controlling for X will be greater than the true causal effect $\mathbb{E}[Y_1]$, and the estimate of the causal effect on Y of intervening to set $A = 0$ controlling for X will be less than the true causal effect $\mathbb{E}[Y_0]$.
ii. If for each U_i, the $U_i - A$ edge (if it exists) is of the opposite sign as the $U_i - Y$ edge (if it exists) then $S_1 \leq \mathbb{E}[Y_1]$ and $S_0 \geq \mathbb{E}[Y_0]$.

If the conditions in (i) are satisfied, then $S_1 - S_0 \geq \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$ and there is positive bias. If the conditions in (ii) are satisfied then $S_1 - S_0 \leq \mathbb{E}[Y_1] - \mathbb{E}[Y_0]$ and there is negative bias. Result 1 can therefore allow the researcher (under the circumstances stated in the result) to determine the sign of the bias, thereby making clear whether, due to lack of control for certain confounding variables, the estimate under consideration is biased towards the null or away from the null. If the estimated risk difference controlling only for X is negative and the conditions in (i) are satisfied, then the estimate of the risk difference controlling only for X is an overestimate of the true causal risk difference. The estimate is thus biased towards the null; if the estimated risk difference controlling only for X is clinically and statistically significant, one could conclude that the true causal effect is also clinically and statistically significant. Similarly, if the estimated risk difference controlling only for X is positive and the conditions in (ii) are satisfied, then the estimate of the risk difference is an underestimate of the true causal risk difference. The estimate is thus again biased towards the null. In such cases, one need not resort to sensitivity analysis techniques to draw conclusions about the presence of a true causal effect, because the direction of the bias is clear and the estimates are conservative. Sensitivity analysis may, however, still be useful in such cases in drawing conclusions about likely upper bounds for the magnitude of the effect. If, however, the estimated risk difference controlling only for X is positive and the conditions in (i) are satisfied, or if it is negative and the conditions in (ii) are satisfied, then the estimate is biased away from the null. In this case, it is not possible to draw conclusions regarding the true causal effect without further sensitivity analysis.

If we return to the example in Figure 1, it follows by Result 1 that if less healthy individuals have a higher probability of receiving the exposure and if less healthy individuals are also more likely to develop the disease, then the estimate of the risk difference not controlling for health status U is in fact conservative for the true causal effect of exposure A on outcome Y.

The use of Result 1 is further illustrated by an example taken from Greenland et al.[2] Consider a study of the effect of antihistamine treatment, denoted by A, on asthma incidence, denoted by Y, among children attending various public schools. Suppose that air pollution levels, denoted by V, is independent of sex, denoted by S, among public school children. Suppose further that sex influences the administration of antihistamine only through its relation to bronchial reactivity, denoted by W, but that sex directly influences asthma risks; that air pollution leads to asthma attacks only through its influence on antihistamine use and bronchial reactivity; and that there are no important confounding variables beyond air pollution, bronchial reactivity, and sex. The causal relationships among these variables are given in Figure 4. If we may furthermore suppose that air pollution has a positive average monotonic effect on bronchial reactivity and on antihistamine use, and that bronchial reactivity has a positive average monotonic effect on antihistamine use and

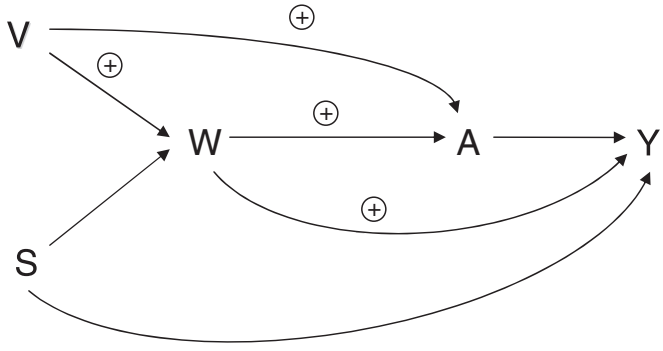

FIGURE 4. Example illustrating the use of Result 1. Y indicates asthma; A, indicates antihistamine treatment; S, indicates sex; V, indicates air pollution; W, indicates bronchial reactivity.

on asthma, then we may also add to the directed edges the positive signs indicated in Figure 4. Suppose that data were available on antihistamine use (yes or no), asthma, air pollution, and sex, but that no data were available for bronchial reactivity. Suppose further that, controlling only for air pollution and sex, antihistamine use was found to lower the risk of asthma. Conditioning on V and S, there is an unblocked backdoor path between A and Y, namely $A - W - Y$. Let $X = \{V, S\}$ and let $U = W$ in the statement of Result 1. The $W - A$ edge and the $W - Y$ edge are of positive sign. Furthermore, the graph in Figure 4 is minimal with respect to A, Y and $X = \{V, S\}$. Since $U = W$, consists of only one variable the condition that the components of U are conditionally independent of one another given X is trivially satisfied. Thus we could conclude from Result 1 that the estimate of the true effect of antihistamine on asthma is to lower asthma risk, ie, that the true causal risk difference, controlling for air pollution, sex, and bronchial reactivity, is negative. This is because the estimate of the risk of asthma when $A = 1$ controlling only for air pollution and sex is an upper bound for $\mathbb{E}[Y_1]$, and the risk of asthma when $A = 0$ controlling only for air pollution and sex is a lower bound for $\mathbb{E}[Y_0]$. Thus if we used observed data and found that $S_1 - S_0 = \sum_{v,s} \mathbb{E}[Y|A = 1, V = v, S = s] P(V = v, S = s) - \sum_{v,s} \mathbb{E}[Y|A = 0, V = v, S = s] P(V = v, S = s) = -0.1$, we could conclude that the true causal effect is such that $\mathbb{E}[Y_1] - \mathbb{E}[Y_0] < -0.1$. Note that if we had data only on A, Y, and S, (ie, if data were unavailable for both V and W) then we could not apply Result 1. This is because if we let $X = S$ and let $U = \{V, W\}$, then the components of U (namely V and W) are not conditionally independent of one another given $X = S$ since V is a cause of W.

When Intuition Fails

There are certain limitations of the application of Result 1 to signed directed acyclic graphs, and contexts in which intuition fails. First, Result 1 requires that the exposure variable under consideration be binary. The result also holds when A is not binary if $A = 1$ is replaced with the maximum value of A and if $A = 0$ is replaced with the minimum value of A. However counterexamples can be constructed to demonstrate that the result cannot be generalized beyond the extreme values of the intervention variable (see counterexample 1 in the online Appendix). Although Result 1 does not hold for intermediate values of the intervention variable, the result can still be useful when the intervention variable A is ordinal or continuous. A nonbinary intervention variable may be dichotomized at various cutpoints. The analysis may proceed with this dichotomized intervention variable, with Result 1 employed to assess the sign of the bias. The analysis may then be repeated at different dichotomization points and conclusions drawn from the resulting analyses.

A second warning is also important: Result 1 applies only to directed acyclic graphs in which the components of U are conditionally independent of one another given X. This

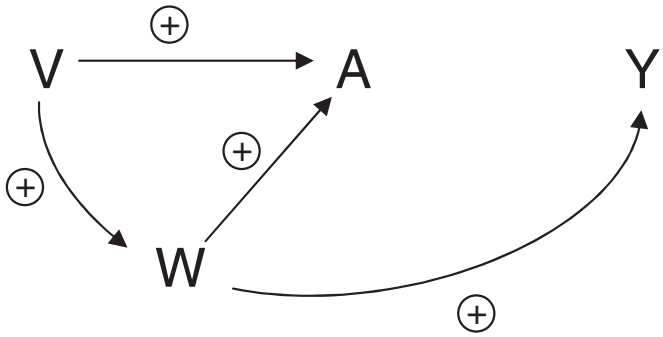

FIGURE 5. Example illustrating that Result 1 may fail for graphs on which the conditional independence condition is not satisfied. Y indicates outcome; A, indicates exposure; V and W, indicate unmeasured confounding variables.

condition is trivially satisfied if U consists of only one variable, ie, if there is only one unmeasured confounder. When U contains more than one variable we can assess the conditional independence condition using the d-separation criterion discussed in the introductory section on directed acyclic graphs. For example, in the graph in Figure 3, the d-separation criterion implies that V is independent of W conditional on X because all paths between V and W are blocked given X. In general, the conditional independence condition will fail whenever 2 components of U are both causes of the same variable in X, or when one component in U is a cause of another component in U. Consider the signed causal directed acyclic graph given in Figure 5. Note if we let $X = \emptyset$, then the graph in Figure 5 is minimal with respect to A and Y. In this example if we let $X = \emptyset$ and $U = \{V, W\}$, then the conditional independence condition of Result 1 will not be satisfied because V is a cause of W. In the online Appendix (counterexample 2) we give a numerical illustration showing that if signs are given to edges for positive and negative average monotonic effects on causal directed acyclic graphs for which the conditional independence condition of Result 1 does not hold, then the conditions in (i) and (ii) are insufficient for drawing the conclusions of Result 1. In the next section we consider assumptions under which the conditional independence condition of Result 1 is not needed. Before moving on we note that the condition that the graph be minimal with respect to A, Y, and X is not strictly necessary; some progress can be made with nonminimal graphs. However, a nonminimal graph has unnecessary variables, the addition of which may violate the conditional independence condition of Result 1.

This brings us to a third warning. Intuitions concerning signed edges can sometimes fail. In particular, examples can be constructed in which an intervention to increase A will increase B on average, and an intervention to increase B will increase C on average, but an intervention to increase A will decrease C on average. The signed causal directed acyclic

FIGURE 6. Example illustrating that positive average monotonic effects are not transitive. A indicates a variable with a positive average monotonic effect on B; B, indicates a variable with a positive average monotonic effect on C; C, indicates outcome.

graph given in Figure 6 and the numerical calculation given in the online Appendix (counterexample 3) illustrate such a case. The relation of a positive average monotonic effect is thus not a transitive relation. In the next section we give stronger conditions—monotonic effects and distributional monotonic effects—under which intuitions concerning signed edges are better preserved. Monotonic effects and distributional monotonic effects do constitute transitive relations. The observation that average monotonic effects are not transitive allows us to give an interpretation to the conditional independence condition of Result 1. As noted above, the conditional independence condition will in general be violated whenever 2 components of U are both causes of the same variable in X, or when one component in U is a cause of another component in U. If 2 components of U are both causes of the same variable in X, the failure of Result 1 may be seen as an instance of conditioning on a common effect or "collider stratification bias."[21,22] If, on the other hand, the conditional independence condition is violated because one component in U is a cause of another component in U, then Result 1 may fail because positive average monotonic effects are not transitive. The conditional independence condition can thus be seen as an assumption that rules out such instances of collider stratification and lack of transitivity.

Monotonic Effects and Distributional Monotonic Effects

As noted above, signs might be given to edges to indicate a variety of relationships. In Appendix 1 we introduce the notions of a monotonic effect and a distributional monotonic effect. The requirements for attributing a monotonic effect or a distributional monotonic effect are considerably stronger than those for a positive average monotonic effect. However, these stronger requirements also allow for stronger conclusions to be drawn. In particular, in the context of monotonic effects or distributional monotonic effects conclusions can be drawn about the direction of unmeasured confounding bias even when there are multiple unmeasured confounding variables that do not satisfy the requirement of Result 1 that the various unmeasured confounding variables are independent of one another conditional on the measured covariates. See Appendix 1 for further details.

Other Measures of Effect

The results in this paper are easily extended to measures of effect other than the causal risk difference. For example, suppose that the outcome Y is binary. If the conditions of Result 1 or Result 2 held and the sign of all unblocked backdoor paths were positive, then for the causal risk ratio one could conclude that $\frac{\sum_x P(Y=1|A=1,X=x)\,P(X=x)}{\sum_x P(Y=1|A=0,X=x)\,P(X=x)} \geq \frac{P(Y_{A=1}=1)}{P(Y_{A=0}=1)}$ and for the causal odds ratio one could conclude that $\frac{\{\sum_x P(Y=1|A=1,X=x)\,P(X=x)\}/\{\sum_x P(Y=0|A=1,X=x)\,P(X=x)\}}{\{\sum_x P(Y=1|A=0,X=x)P(X=x)\}/\{\sum_x P(Y=0|A=0,X=x)\,P(X=x)\}} \geq \frac{P(Y_{A=1}=1)/P(Y_{A=1}=0)}{P(Y_{A=0}=1)/P(Y_{A=0}=0)}$. If the sign of all unblocked backdoor paths were negative, then the direction of the inequalities would be reversed.

DISCUSSION

We have formalized the conditions under which signs can be added to the edges of a causal directed acyclic graph. We have also given results that formalize conclusions about the direction of the bias that are often drawn intuitively, and we have described the cases in which such intuition may fail. Signs can be added to edges when intervening on the parent node increases on average the value of the child node regardless of the values of the other parents. These signs can be used to draw conclusions about the direction of the bias of unmeasured confounding (see Result 1). If only one unmeasured confounding variable is present, it is relatively easy to draw conclusions about the direction of the bias. When higher values of the unmeasured confounding variable increase on average both the exposure and the outcome, then there will be positive bias. If higher values of the unmeasured confounding variable increase on average either the exposure or the outcome and decrease on average the other, then there will be negative bias. If the estimate without controlling for the unmeasured confounding variable is positive and the direction of the bias is negative, then we could conclude that the estimate without controlling for unmeasured confounding is biased towards the null and thus conservative. In such cases we could conclude the presence of a true causal effect without using sensitivity analysis techniques, although sensitivity analysis might still be useful in giving an upper bound on the magnitude of the effect. Similarly, if the estimate without controlling for the unmeasured confounding variable is negative and the direction of the bias is positive, then we could conclude that the estimate without controlling for unmeasured confounding is again biased towards the null and thus conservative.

If there is more than one unmeasured confounding variable, the same principles apply but somewhat stronger assumptions are needed. Specifically, if there are multiple

unmeasured confounding variables, then we need to impose some restrictions on the relationships between the different unmeasured confounding variables. Specifically it must be that no unmeasured confounding variable is the cause of another unmeasured confounding variable, and it must also be the case that, if there are measured covariates for which control is being made, then no 2 unmeasured confounding variables can be causes of the same measured covariate. If there are multiple unmeasured confounding variables and these conditions are not met, counterintuitive results can sometimes occur. Even if these conditions are not met, progress can still sometimes be made concerning the direction of the bias but stronger notions of monotonicity are then needed (discussed in Appendix 1).

To use the results in this paper, some knowledge of the relationship between the unmeasured confounder and the exposure and between the unmeasured confounder and the outcome are necessary, namely whether the unmeasured confounder increases or decreases on average the exposure and the outcome. When such knowledge is available, our results can be useful in drawing conclusions about the direction of the bias that results from unmeasured confounding. The results can thereby be useful in drawing conclusions about the presence of a true causal effect even in the presence of unmeasured confounding.

APPENDIX 1

Monotonic Effects and Distributional Monotonic Effects

Here, we introduce the notions of a monotonic effect and a distributional monotonic effect. The requirements for attributing a positive monotonic effect are considerably stronger than those for a positive average monotonic effect. The definition of a monotonic effect essentially requires that some intervention S either increases or decreases some other variable T not merely on average over the entire population but rather for every individual in that population regardless of the inventions made on the other parents of T. More formally, if a variable S is a parent of some variable T and Q is the set of parents of T other than S then we will say that S has a *positive monotonic effect* on T if for all individuals ω in the population and all values of q, $T_{s_1,q}(\omega) \geq T_{s_0,q}(\omega)$ whenever $s_1 \geq s_0$ where $T_{s,q}(\omega)$ is the counterfactual value for individual ω intervening to set $S = s$ and $Q = q$. We will say that S has a *negative monotonic effect* on T if for all individuals ω in the population and all values of q, $T_{s_1,q}(\omega) \leq T_{s_0,q}(\omega)$ whenever $s_1 \geq s_0$. The requirements for the attribution of a monotonic effect are thus considerable. However, whenever a particular intervention is always beneficial or neutral for all individuals with respect to a particular outcome, one will be able to attribute a positive monotonic effect; whenever the intervention is always harmful or neutral for all individuals with respect to a particular outcome, one will be able to attribute a negative monotonic effect. Examples of monotonic effects might include the effect of smoking on lung cancer or the effect of certain environmental exposures or genes on particular outcomes. However, because for any individual we observe the counterfactual outcome only under one particular value of the intervention variable, the presence of a monotonic effect is not identifiable and we must thus rely on substantive knowledge of the problem under consideration to attribute a monotonic effect.

The requirements for a distributional positive monotonic effect are between those for a positive monotonic effect and those for a positive average monotonic effect. The presence of a distributional monotonic requires that for all t a higher value of S makes the probability of event $\{T \geq t\}$ over the whole population more likely or as likely regardless of the value the parents of T other than S. More formally, suppose that variable S is a parent of some variable T and let Q denote the parents of T other than S. We say that S has a *positive distributional monotonic effect* (or a *weak positive monotonic effect*[14]) on T if the survivor function $P(T \geq t|S = s, Q = q)$ is such that whenever $s_1 \geq s_0$ we have $P(T \geq t|S = s_1, Q = q) \geq P(T \geq t|S = s_0, Q = q)$ for all t and all q; the variable S is said to have a *negative distributional monotonic effect* on T if whenever $s_1 \geq s_0$ we have $P(T \geq t|S = s_1, Q = q) \leq P(T \geq t|S = s_0, Q = q)$ for all t and all q.

The presence of a distributional monotonic effect is a substantially less stringent condition than that of a monotonic effect. If intervening to increase S led to a decrease in T for only a single individual the strong conditions for a monotonic effect would fail. The less stringent conditions required for attributing a distributional monotonic effect circumvents this difficulty. Consider, for example, an analysis comparing the effect on thyroid cancer of no radiation exposure to a high level of radiation exposure. For most individuals the exposure to a high level of radiation will increase the likelihood of developing thyroid cancer. However, exposure to a high level of radiation may, for a few individuals, destroy already existing thyroid cancer cells and thereby prevent the cancer's development. Within joint strata of particular sets of background variables on a causal directed acyclic graph, the exposure to radiation will increase the overall likelihood of thyroid cancer but it may not do so for every individual in the population. In such a scenario the high level of radiation exposure would not have a monotonic effect on the development of thyroid cancer but it would have a distributional monotonic effect.

It can be shown that the presence of a positive monotonic effect implies the presence of a positive distributional monotonic effect and that the presence of a positive distributional monotonic effect implies the presence of a positive average monotonic effect. In the case of a binary outcome T, a positive distributional monotonic effect and a positive average monotonic effect are equivalent. Unlike positive average mono-

tonic effects, positive monotonic effects and positive distributional monotonic effects are transitive. Theorems concerning the transitivity of monotonic effects and distributional monotonic effects have been given in related work.[14]

When signs are given to edges in the presence of monotonic effects or distributional monotonic effects we can draw conclusions about the sign of the bias even when the conditional independence condition of Result 1 does not hold. Thus, although the requirements for monotonic effects and distributional monotonic effects are considerable, they do allow the researcher to draw conclusions in a greater number of contexts. The following Result was proved by Vander-Weele and Robins.[14]

Result 2. Suppose that for some binary intervention A and some outcome Y, some set X of nondescendents of A does not block all backdoor paths from A to Y but does not open any backdoor paths from A to Y which were blocked without conditioning on X. Suppose also that X has no ancestors outside of the set X. Suppose further that signs are given to edges only for monotonic effects or distributional monotonic effects. Let $S_a = \sum_x \mathbb{E}[Y|A = a, X = x]P(X = x)$. If all unblocked backdoor paths from A to Y are of positive sign then $S_1 \geq \mathbb{E}[Y_1]$ and $S_0 \leq \mathbb{E}[Y_1]$. If all unblocked backdoor paths from A to Y are of negative sign then $S_1 \leq \mathbb{E}[Y_1]$ and $S_0 \geq \mathbb{E}[Y_1]$.

Let us return to the causal directed acyclic graph given in Figure 4. Suppose that data were only available for A, Y, and S but that the signed edges on Figure 4 represented not merely average monotonic effects but distributional monotonic effects. Conditioning only on S, there are 2 unblocked backdoor paths between A and Y: $A - W - Y$ and $A - V - W - Y$. The sign of both of these unblocked backdoor paths from A to Y are positive. Note that S has no ancestors. We could conclude from Result 2 that the estimate controlling only for S is biased towards the null and thus conservative for the true causal effect of A on Y. As noted above, to draw this conclusion when data is only available on S we cannot use Result 1. Average monotonic effects are not sufficient here because the conditional independence condition is not satisfied; in this case we need distributional monotonic effects so that we can apply Result 2.

APPENDIX 2

Proof of Result 1

We will use that notation $A \amalg B|C$ to denote that A is independent of B conditional on C. Using this notation, the causal effect of A on Y is said to be unconfounded given Z if $Y_a \amalg A|Z$. Note that Pearl's backdoor paths criterion thus can be stated as follows: if a set Z of nondescendents of A blocks all backdoor paths from A to Y then $Y_a \amalg A|Z$. In the proof we will make use of the following result concerning potential outcomes[20]: Suppose that A is binary and that (1)$Y_a \amalg A|\{X, U\}$ for $a = 0, 1$, (2) $\mathbb{E}[Y|A = a, X = x, U = u]$ is nondecreasing in u for all a and x, (3) $\mathbb{E}[A|X = x, U = u]$ is nondecreasing in u for all x and (4) if U is multivariate then the components of U are conditionally independent given X then $\sum_x \mathbb{E}[Y|A = 1, X = x]P(X = x) \geq \mathbb{E}[Y_1]$ and $\sum_x \mathbb{E}[Y|A = 0, X = x]P(X = x) \leq \mathbb{E}[Y_0]$. We use this result to prove the result given in the present paper. Consider the case in which the conditions in part (i) of Result 1 are satisfied; when the conditions in (ii) are satisfied the proof is analogous. Since X and U include all nondescendents of A on the graph, X and U must block all backdoor paths from A to Y and from Pearl's backdoor path criterion it follows that (1) holds. Condition (4) holds by assumption. We will now show that conditions (2) and (3) hold. For every node U_i such that the edges $U_i - A$ and $U_i - Y$, if they exist, are of negative sign, we may replace U_i on the graph with its negation U_i so that the edges into A and Y are of positive sign. We may thus assume without loss of generality that if the conditions in (i) hold then every edge from each U_i to A and to Y, if they exist, are of positive sign. Let pa_Y and pa_A denote the parents of Y and A respectively. Since X and U contain all the variables on the graph other than A and Y we have that $pa_Y \subseteq A \cup X \cup U$ and $pa_A \subseteq X \cup U$. Thus $\mathbb{E}[Y|A = a, X = x, U = u] = \mathbb{E}[Y|pa_Y]$ and since every edge from U_i to Y is of positive sign we have that $\mathbb{E}[Y|A = a, X = x, U = u] = \mathbb{E}[Y|pa_Y]$ is nondecreasing in u by the definition of a positive average monotonic effect. Similarly, $\mathbb{E}[A|X = x, U = u] = \mathbb{E}[A|pa_A]$ and since every edge from U_i to A is of positive sign we have that $\mathbb{E}[A|X = x, U = u] = \mathbb{E}[A|pa_A]$ is nondecreasing in u. Conditions (2) and (3) thus hold and the conclusion follows.

REFERENCES

1. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82: 669–688.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
3. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology*. 2001;12:313–320.
4. Hernán MA, Hernández-Díaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155:176–184.
5. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: Recent evidence and a discussion of some questions. *J Natl Cancer Inst*. 1959;22:173–203.
6. Manski C. Nonparametric bounds on treatment effects. *Am Econ Rev*. 1990;80:319–323.
7. MacLehose RF, Kaufman S, Kaufman JS, et al. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*. 2005;16:548–555.
8. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J Roy Stat Soc Ser B*. 1983;45:212–218.
9. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounding in observational studies. *Biometrics*. 1998;54:948–963.
10. Hernán MA, Robins JM. Letter to the editor of Biometrics. *Biometrics*. 1999;55:1316–1317.

11. Brumback BA, Hernán MA, Haneuse SJPA, et al. Sensitivity analyses for unmeasured confounding assuming a average monotonic structural model for repeated measures. *Stat Med*. 2004;23:749–767.
12. Greenland S. Multiple-bias modelling for analysis of observational data. *J Roy Stat Soc Ser A*. 2005;168:267–306.
13. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25:1107–1116.
14. VanderWeele TJ, Robins JM. Signed directed acyclic graphs for causal inference. In: VanderWeele TJ. Contributions to the theory of causal directed acyclic graphs [PhD thesis]. Cambridge, MA: Harvard University; 2006:1:–42.
15. Pearl J. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press; 2000.
16. Spirtes P, Glymour C, Scheines R. *Causation, Prediction and Search.* New York: Springer-Verlag; 1993.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
18. Robins JM. A new approach to causal inference in mortality studies with sustained exposure period—application to control of the healthy worker survivor effect. *Math Modelling*. 1986;7:1393–1512.
19. Robins JM. Addendum to a new approach to causal inference in mortality studies with sustained exposure period—application to control of the healthy worker survivor effect. *Comput Math Appl*. 1987;14: 923–945.
20. VanderWeele TJ. The sign of the bias of unmeasured confounding. *Biometrics*. January 4, 2008 [Epub ahead of print].
21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
22. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.