Directed acyclic graphs with edge-specific bounds

BY TYLER J. VANDERWEELE

*Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue,
Boston, Massachusetts 02115, U.S.A.*

tvanderw@hsph.harvard.edu

AND ZHIQIANG TAN

*Department of Statistics, Rutgers, The State University of New Jersey, 110 Frelinghuysen Road,
Piscataway, New Jersey 08854, U.S.A.*

ztan@stat.rutgers.edu

SUMMARY

We give a definition of a bounded edge within the causal directed acyclic graph framework. A bounded edge generalizes the notion of a signed edge and is defined in terms of bounds on a ratio of survivor probabilities. We derive rules concerning the propagation of bounds. Bounds on causal effects in the presence of unmeasured confounding are also derived using bounds related to specific edges on a graph. We illustrate the theory developed by an example concerning estimating the effect of antihistamine treatment on asthma in the presence of unmeasured confounding.

Some key words: Bayesian network; Bound; Causal inference; Confounding; Directed acyclic graph.

1. INTRODUCTION

Building on [Wellman \(1990\)](#), [VanderWeele & Robins \(2009, 2010\)](#) developed theory for signed causal directed acyclic graphs and derived results that relate signed edges to causal effects, to covariance amongst variables, and to the sign of the bias that results when unmeasured confounding is present. Signed edges amount to statements about ratios of survivor probabilities, that the ratios are bounded either between 0 and 1 for negative edges or between 1 and ∞ for positive edges, but in certain cases these bounds may be too restrictive. If, for example, the bounds for the ratio are of the form (a, b) with $a < 1 < b$ and thus include 1 rather than being bounded above or below by 1, a signed edge cannot be assigned. In other cases, the ratios might be known to lie in ranges of the form $(c, 1)$ or $(1, 1/c)$, where $0 < c < 1$. In this paper, we generalize the definitions for a weak monotonic effect and a signed edge given by [VanderWeele & Robins \(2010\)](#) to the case of bounded edges and derive results concerning these bounded edges.

2. CAUSAL DIRECTED ACYCLIC GRAPHS AND SIGNED EDGES

A directed graph ([Spirtes et al., 1993](#); [Pearl, 1995, 2000](#); [Dawid, 2002](#)) consists of a set of nodes and directed edges amongst nodes. A path is a sequence of distinct nodes connected by edges regardless of arrowhead direction; a directed path is a path which follows the edges in the direction indicated by the graph's arrows. A directed acyclic graph is a directed graph in which no node has a directed path back to itself. The nodes with directed edges into a node

A are said to be the parents of A ; the nodes into which there are directed edges from A are said to be its children. We say that node A is an ancestor of node B if there is a directed path from A to B , and B is then called a descendant of A . A node is called a collider for a particular path if both the preceding and subsequent nodes on the path have directed edges going into it. A path between two nodes, A and B , is said to be blocked given some set of nodes C if either there is a variable in C on the path that is not a collider for the path or if there is a collider on the path such that neither the collider itself nor any of its descendants are in C . For disjoint sets of nodes A , B and C , we say that A and B are d -separated given C if every path from any node in A to any node in B is blocked given C . Directed acyclic graphs are sometimes used as statistical models to encode independence relationships amongst variables represented by the nodes on the graph (Lauritzen, 1996). We will use the notation $A \perp\!\!\!\perp B \mid C$ to denote that A is conditionally independent of B given C . The variables corresponding to the nodes on a graph are said to satisfy the global Markov property for the directed acyclic graph if for any disjoint sets of nodes A , B , C we have that $A \perp\!\!\!\perp B \mid C$ whenever A and B are d -separated given C .

Directed acyclic graphs can be interpreted as representing causal relationships (Spirites et al., 1993; Pearl, 1995, 2000; Dawid, 2002). Let Y_a denote the counterfactual value of Y under an intervention to set A to a . Pearl (1995) defined a causal directed acyclic graph as a directed acyclic graph with nodes $(X_1, \dots, X_n)$ corresponding to variables such that each variable X_i is given by its nonparametric structural equation $X_i = f_i(\text{pa}_i, \epsilon_i)$ where pa_i are the parents of X_i on the graph and the ϵ_i are mutually independent. These structural equations generalize the path analysis and linear structural equation models (Pearl, 1995, 2000) developed by Wright (1921) in the genetics literature and Haavelmo (1943) in the econometrics literature. The structural equations encode counterfactual relationships amongst the variables represented on the graph, themselves representing one-step ahead counterfactuals, with other counterfactuals given by recursive substitution. On a causal directed acyclic graph, a node C is said to be a common cause of A and Y if there exists a directed path from C to Y not through A and a directed path from C to A not through Y . The requirement that the ϵ_i be mutually independent is essentially a requirement that there is no variable absent from the graph which, if included on the graph, would be a parent of two or more variables (Pearl, 1995, 2000). A causal directed acyclic graph defined by nonparametric structural equations satisfies the global Markov property as stated above (cf. Verma & Pearl, 1988; Geiger et al., 1990; Lauritzen et al., 1990; Pearl, 2000). We will say that $(V_1, \dots, V_n)$ constitutes an ordered list if $i < j$ implies that V_i is not a descendant of V_j . On some graphs there will be more than one possible ordering. The results below will apply provided that there is some ordering under which the conditions of the propositions and theorems are satisfied. We will use $\bar{V}_k$ to denote $(V_1, \dots, V_k)$, with $\bar{V}_0 = \emptyset$. For further discussion of the causal interpretation of directed acyclic graphs see Spirites et al. (1993), Pearl (1995, 2000), Dawid (2002) and Robins (2003).

The directed acyclic graph causal framework has proved to be particularly useful in determining whether conditioning on a given set of variables, or none at all, is sufficient to control for confounding. The most important result in this regard is the back-door path criterion (Pearl, 1995). A back-door path from some node A to another node Y is a path into Y which begins with a directed edge into A ; a front-door path from A to Y is a path into Y which begins with a directed edge emanating from A . Pearl (1995) showed that for intervention variable A and outcome Y , if a set of variables X is such that no variable in X is a descendant of A and such that X blocks all back-door paths from A to Y then $Y_a \perp\!\!\!\perp A \mid X$, so that conditioning on X suffices to control for confounding for the estimation of the causal effect of A on Y . We will use the idea of a back-door path throughout the paper.

Some of the results below only make reference to conditional probability distributions rather than counterfactuals and thus do not require a causal interpretation of directed acyclic graphs. Nevertheless, we believe the results in this paper will be of greatest interest in drawing inferences concerning causal effects.

If v is a vector, then we will say some function $f(v)$ is nondecreasing in v if it is nondecreasing in each component of v . If A is a parent of Y , we will use pa_Y^A to denote the parents of Y other than A . Throughout, we assume that the conditional distributions used are well defined (e.g., Billingsley, 1995, § 33). For example, we assume that $\text{pr}(Y \leq y \mid A = a, \text{pa}_Y^A)$ as a function of y for fixed a and pa_Y^A is a proper cumulative distribution function. By convention, we interpret the fraction $y/0$ as infinite for any $y > 0$.

VanderWeele & Robins (2009, 2010) gave the following definitions for a weak monotonic effect and a signed edge on a causal directed acyclic graph (cf. Wellman, 1990).

DEFINITION 1. *We say that A has a weak positive monotonic effect on Y if the survivor function $S(y \mid a, \text{pa}_Y^A) = \text{pr}(Y > y \mid A = a, \text{pa}_Y^A)$ is such that whenever $a_1 \geq a_0$ we have $S(y \mid a_1, \text{pa}_Y^A) \geq S(y \mid a_0, \text{pa}_Y^A)$ for all y and all pa_Y^A , and a weak negative monotonic effect if whenever $a_1 \geq a_0$ we have $S(y \mid a_1, \text{pa}_Y^A) \leq S(y \mid a_0, \text{pa}_Y^A)$ for all y and all pa_Y^A .*

Corresponding to the notion of a weak monotonic effect is that of a signed edge.

DEFINITION 2. *An edge on a causal directed acyclic graph from A to Y is said to be of positive or negative sign if, respectively, A has a weak positive or negative monotonic effect on Y ; otherwise, it is without sign. The sign of a path is the product of the signs of the edges that constitute that path; the path is without sign if one of the path's edges is without sign.*

The definition of a weak positive monotonic effect requires that for all $a_1 \geq a_0$ such that at least one of $S(y \mid a_1, \text{pa}_Y^A)$ or $S(y \mid a_0, \text{pa}_Y^A)$ is nonzero, $1 \leq S(y \mid a_1, \text{pa}_Y^A)/S(y \mid a_0, \text{pa}_Y^A) \leq \infty$; a weak negative monotonic effect requires that for all $a_1 \geq a_0$ such that at least one of $S(y \mid a_1, \text{pa}_Y^A)$ or $S(y \mid a_0, \text{pa}_Y^A)$ is nonzero, $0 \leq S(y \mid a_1, \text{pa}_Y^A)/S(y \mid a_0, \text{pa}_Y^A) \leq 1$ whenever $a_1 \geq a_0$. The restriction that such inequalities only need to be satisfied when either the numerator or the denominator is nonzero will be assumed to apply to all ratios throughout and this condition will not be repeatedly stated.

3. BOUNDED EDGES AND THE PROPAGATION OF BOUNDS

We now give a definition for a bounded edge that generalizes the notion of a signed edge.

DEFINITION 3. *For some node Y with parent A we will say that the $A \rightarrow Y$ edge is stochastically bounded by (Λ^-, Λ^+) if*

$$\Lambda^- \leq \frac{\text{pr}(Y > y \mid A = a_1, \text{pa}_Y^A)}{\text{pr}(Y > y \mid A = a_0, \text{pa}_Y^A)} \leq \Lambda^+,$$

for all $a_1 > a_0$, y and pa_Y^A . If the $A \rightarrow Y$ edge is stochastically bounded by (Λ^-, Λ^+) , then we write $\Theta(A, Y) = (\Lambda^-, \Lambda^+)$ and place the ordered set (Λ^-, Λ^+) on the $A \rightarrow Y$ edge of the directed acyclic graph. If there is no edge from A to Y on the graph then we define $\Theta(A, Y) = (1, 1)$.

We will refer to a directed acyclic graph with bounds placed on the edges of the graph as a bounded directed acyclic graph. For any edge we must have that $0 \leq \Lambda^- < 1$ and $1 < \Lambda^+ \leq \infty$;

the bounds $(0, \infty)$ can be placed on any edge, equivalent to an edge without bounds. By letting $y \rightarrow -\infty$ in the ratio in Definition 3, we see that the interval (Λ^-, Λ^+) must contain 1. If $\Theta(A, Y) = (1, b)$, then A has a weak positive monotonic effect on Y and the $A \rightarrow Y$ edge will be of positive sign; if $\Theta(A, Y) = (a, 1)$, then the $A \rightarrow Y$ edge will be of negative sign. If Y is binary, then $\Theta(A, Y) = (1, \Lambda^+)$ simply requires that

$$1 \leq \frac{\text{pr}(Y = 1 \mid A = a_1, \text{pa}_Y^A)}{\text{pr}(Y = 1 \mid A = a_0, \text{pa}_Y^A)} \leq \Lambda^+,$$

for all $a_1 > a_0$, pa_Y^A . Definitions 1–3 all apply to any statistical graphical model that satisfies the global Markov property for a directed acyclic graph.

To develop theory on the propagation of bounds on a directed acyclic graph, we rely on the following proposition. Its proof and those of other results are given in the Appendix.

PROPOSITION 1. *For fixed a_0, a_1 and q , assume that the following conditions hold.*

(a) *For $i = 1, \dots, n$,*

$$\Lambda_i^- \leq \frac{\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_1, Q = q)}{\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_0, Q = q)} \leq \Lambda_i^+,$$

for all v_i and $\bar{v}_{i-1}$;

(b) *$\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_1, Q = q)$ and $\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_0, Q = q)$ are nondecreasing in $\bar{v}_{i-1}$ ($i = 2, \dots, n$); and*

(c) *$G(\bar{v}_n)$ is nondecreasing in $\bar{v}_n$.*

Then

$$\prod_{i=1}^n \Lambda_i^- \leq \frac{\text{pr}\{G(\bar{V}_n) > g \mid A = a_1, Q = q\}}{\text{pr}\{G(\bar{V}_n) > g \mid A = a_0, Q = q\}} \leq \prod_{i=1}^n \Lambda_i^+, \quad (1)$$

for all g .

Proposition 1 does not require reference to a directed acyclic graph; the conclusion holds for any ordered sequence $V_1, \dots, V_n$ that satisfies (a) and (b). The next theorem essentially says that if we can find some set X that blocks all back-door paths from a node A to a node Y , and if certain directed paths into Y are signed, then bounds for the effect of A on Y can be derived from the bounds on the edges emanating from A .

THEOREM 1. *Suppose that A is an ancestor of Y and that some set X of nondescendants of A blocks all back-door paths from A to Y . Let $(V_1, \dots, V_n)$ be an ordered list of all nodes on directed paths from A to Y and let $V_{n+1} = Y$. We may denote the bounds on the edges from A to V_i by $\Theta(A, V_i) = (\Lambda_i^-, \Lambda_i^+)$ for $i = 1, \dots, n+1$. If for all i such that there is an edge from A to V_i all directed paths from V_i to Y are of positive sign, then for all $G(\bar{v}_{n+1})$ nondecreasing in $\bar{v}_{n+1}$,*

$$\prod_{i=1}^{n+1} \Lambda_i^- \leq \frac{\text{pr}\{G(\bar{V}_{n+1}) > g \mid A = a_1, X = x\}}{\text{pr}\{G(\bar{V}_{n+1}) > g \mid A = a_0, X = x\}} \leq \prod_{i=1}^{n+1} \Lambda_i^+,$$

for $a_1 > a_0$ and all g .

Theorem 1 allows for settings in which there is some node V_j on a directed path from A to Y such that there is no edge from A to V_j ; in this case $\Lambda_j^- = \Lambda_j^+ = 1$. If for some i , all directed

Fig. 1. Example of the propagation of bounds.

paths from V_i to Y are of negative, rather than positive, sign and $G(\bar{v}_{n+1})$ is nonincreasing, rather than nondecreasing, in v_i then, the result could still be applied by replacing V_i with $-V_i$ so that all directed paths from $-V_i$ to Y are of positive sign. The bounds Λ_i^- , Λ_i^+ will then have to be specified so that $\Theta(A, -V_i) = (\Lambda_i^-, \Lambda_i^+)$. If $\Theta(A, V_i)$ is of form $(a, 1)$, then $\Theta(A, -V_i)$ is of form $(1, b)$, but b is not in general $1/a$.

VanderWeele & Robins (2009) showed that if X is a set of nondescendants of A that blocked all back-door paths from A to Y and if all directed paths from A to Y are of positive sign then, $\text{pr}(Y > y | a, x)$ is nondecreasing in a . If in Theorem 1 $\Theta(A, V_i) = (\Lambda_i^-, \Lambda_i^+)$ is of the form $(1, \infty)$ for all i , then, equivalently, under the assumptions of Theorem 1, all directed paths from A to Y are positive. This special case of Theorem 1 is still a generalization of the aforementioned result of VanderWeele & Robins (2009). The reason is that under these assumptions, Theorem 1 would allow one to conclude that $\text{pr}\{G(\bar{v}_{n+1}) > g | a, x\}$ was nondecreasing in a for any choice of the function $G(\bar{v}_{n+1})$ nondecreasing in $\bar{v}_{n+1}$, rather than simply for the special choice of $G(\bar{v}_{n+1})$ as $G(\bar{v}_{n+1}) = y$.

If $G(\bar{v}_{n+1})$ is taken as $G(\bar{v}_{n+1}) = y$, then by Pearl's back-door path adjustment theorem (Pearl, 1995), it follows immediately from Theorem 1 that

$$\prod_{i=1}^{n+1} \Lambda_i^- \leq \frac{\text{pr}(Y_{a_1} > y | X = x)}{\text{pr}(Y_{a_0} > y | X = x)} \leq \prod_{i=1}^{n+1} \Lambda_i^+.$$

We illustrate the use of Theorem 1 in the following example.

Example 1. Consider the bounded directed acyclic graph given in Fig. 1. All directed paths from V_1 to Y are of positive sign and all directed paths from V_2 to Y are of positive sign. The path consisting of the edge $V_1 \rightarrow Y$ is of positive sign since $\Theta(V_1, Y) = (1, 4)$; the path $V_1 \rightarrow V_3 \rightarrow Y$ is of positive sign since the edge $V_1 \rightarrow V_3$ is of negative sign and the edge $V_3 \rightarrow Y$ is of negative sign and thus the product of the signs of these edges is positive. Finally, the path consisting of the edge $V_2 \rightarrow Y$ is also of positive sign. There are edges emanating from A into V_1 , V_2 and Y with bounds $\Theta(A, V_1) = (1, 3)$, $\Theta(A, V_2) = (1, 2)$, $\Theta(A, Y) = (2/3, 2)$, and by Theorem 1, we have for all c and all $a_1 > a_0$ that $2/3 = (1)(1)(2/3) \leq \text{pr}(Y > y | A = a_1, C = c) / \text{pr}(Y > y | A = a_0, C = c) \leq (3)(2)(2) = 12$ since C blocks all back-door paths from A to Y ; similarly, $2/3 \leq \text{pr}(Y > y | A = a_1, X = x) / \text{pr}(Y > y | A = a_0, X = x) \leq 12$ since X also blocks all back-door paths from A to Y .

Others have tried to generalize the notion of a signed edge in order to account for additional information or numeric bounds. Parsons (1995) provides a set of possible axiomatic rules to

govern the propagation of influences on networks that can be strongly or weakly positive or negative, rather than simply positive or negative. Parsons thus extended Wellman's qualitative influence to the notion of categorical influence. [Renooij & van der Gaag \(2008\)](#) have recently further developed Parsons' approach, but neither provide the generality of our notion of a bounded edge. A more related approach is that of [Liu & Wellman \(1998, 2004\)](#) who provide bounds for a cumulative distribution function, not by bounding the ratios of survivor probabilities but by postulating that certain cumulative distribution functions stochastically dominate those in fact governing various signed edges on a graph. Their results also differ from ours in another way: while in Theorem 1 bounds of the form $\text{pr}(Y > y \mid A = a_1, X = x) / \text{pr}(Y > y \mid A = a_0, X = x)$ are obtained by information on bounded edges emanating from A , their results require bounds on the cumulative distribution functions corresponding to edges pointing into Y . Additional research might consider if further inferences concerning bounds would be possible if their results were combined with ours.

4. BOUNDS FOR CAUSAL EFFECTS IN THE PRESENCE OF UNMEASURED CONFOUNDING

The following result allows us to give bounds for a causal effect in the presence of unmeasured confounding.

THEOREM 2. *Suppose that for some variable A and some nonnegative outcome Y , the set $X = C \cup U$ of nondescendants of A blocks all back-door paths from A to Y , where C consists of measured covariates and U unmeasured covariates. Let $S^a = \sum_c E(Y \mid a, c) \text{pr}(c)$. If for some u' and some $\Lambda_L, \Lambda_H > 0$,*

$$\Lambda_L E(Y \mid a, u', c) \leq E(Y \mid a, u, c) \leq \Lambda_H E(Y \mid a, u', c), \quad (2)$$

for all a, c and u , then

$$\frac{\Lambda_L}{\Lambda_H} S^a \leq E(Y_a) \leq \frac{\Lambda_H}{\Lambda_L} S^a.$$

For (2) to hold for $u = u'$, the interval (Λ_L, Λ_H) must contain 1.

COROLLARY 1. *Under the assumptions of Theorem 2 we have*

$$\frac{\Lambda_L}{\Lambda_H} S^{a_1} - \frac{\Lambda_H}{\Lambda_L} S^{a_0} \leq E(Y_{a_1}) - E(Y_{a_0}) \leq \frac{\Lambda_H}{\Lambda_L} S^{a_1} - \frac{\Lambda_L}{\Lambda_H} S^{a_0} \quad (3)$$

and

$$\left(\frac{\Lambda_L}{\Lambda_H} \right)^2 \frac{S^{a_1}}{S^{a_0}} \leq \frac{E(Y_{a_1})}{E(Y_{a_0})} \leq \left(\frac{\Lambda_H}{\Lambda_L} \right)^2 \frac{S^{a_1}}{S^{a_0}}.$$

In Theorem 2, if Λ_L and Λ_H are known a priori from subject matter knowledge then $S^a = \sum_c E(Y \mid a, c) \text{pr}(c)$ can be estimated from data and thus, by using Corollary 1, one can obtain bounds for the causal effect, $E(Y_{a_1}) - E(Y_{a_0})$ on the additive scale, or $E(Y_{a_1})/E(Y_{a_0})$ on the multiplicative scale. If Λ_L and Λ_H are unknown, Corollary 1 could still be used in sensitivity analysis ([Cornfield et al., 1959](#)) by varying Λ_L and Λ_H . The assumptions required for the use of these results in sensitivity analysis are much weaker than those required for other techniques (e.g., [Lin et al., 1998](#)). Bounded edges in conjunction with Theorem 1 can also be used to yield bounds for inequality (2) by means of the following proposition.

Fig. 2. Example concerning bounds for the effect of antihistamine use on asthma in the presence of unmeasured confounding.

PROPOSITION 2. *For nonnegative Y , if for all y , $\text{pr}(Y > y | u, l) \leq \Lambda \text{pr}(Y > y | u', l)$, then $E(Y | u, l) \leq \Lambda E(Y | u', l)$.*

If in Theorem 2, U is univariate and (A, C) contain all parents of Y other than U then the bounds for the $U \rightarrow Y$ edge, say $\Theta(U, Y) = (\Lambda_L, \Lambda_H)$, will, by Proposition 2 with $L = (A, C)$, be bounds that satisfy (2) provided there is some minimum value u' of U . However, although the application of Proposition 2 could be used to draw conclusions about the inequality in (2), (2) is in fact weaker than what is required for a bounded edge and Theorem 2 can thus be employed more generally. For example, suppose (A, C) are nondescendants of U and block all back-door paths from U to Y , although U lies on a back-door path from A to Y . Then the bounds relating U and Y from Theorem 1, with U and (A, C) taking the roles of A and X , can be used to give bounds that satisfy (2) by Proposition 2. In the next section, we apply Theorem 2 to a problem concerning the effect of antihistamine treatment on asthma in the presence of unmeasured confounding.

5. APPLICATION AND FURTHER DISCUSSION

We present an example adapted from [Greenland et al. \(1999\)](#) and discussed by [VanderWeele et al. \(2008\)](#). Using bounds for edges rather than signed edges allows us to derive bounds for the causal effect under fewer assumptions.

Example 2. Consider a hypothetical study of the relation of antihistamine treatment, denoted by A , and asthma incidence, denoted by Y , among first-grade children attending public schools. Suppose that air pollution level, denoted by W , is independent of sex, denoted by C . Suppose further that sex influences the administration of antihistamine only through its relation to bronchial reactivity, denoted by U , but directly influences asthma risk; suppose also that air pollution leads to asthma attacks only through its influence on antihistamine use and bronchial reactivity; and that there are no important confounders beyond air pollution, bronchial reactivity and sex. The causal relationships amongst these variables are then those given in Fig. 2.

Under the assumptions given above, by Pearl's back-door path criterion conditioning on C and U suffices to control for confounding; conditioning on C , U and W or on U and W also suffices. If data were only available for antihistamine use A , asthma Y and sex C , then we could not produce valid estimates of the causal effect of A on Y because controlling only for C does not suffice to control for confounding. Suppose now that, for the purposes of this study, asthma and bronchial reactivity can be considered binary, comparing high versus low, and that $1 \leq \text{pr}(Y = 1 | a, U = 1, c) / \text{pr}(Y = 1 | a, U = 0, c) \leq 2$ for all a, c . In other words, high bronchial reactivity increases the likelihood of asthma by a factor somewhere between 1 and 2

for all levels of antihistamine treatment for both males and females. Suppose that in the analysis of the available data it was found that $S^1 = \sum_c E(Y | A = 1, C = c) \text{pr}(C = c) = 0.06$ and $S^0 = \sum_c E(Y | A = 0, C = c) \text{pr}(C = c) = 0.25$. From Corollary 1 we have that

$$\frac{1}{2}S^1 - 2S^0 \leq E(Y_{a=1}) - E(Y_{a=0}) \leq 2S^1 - \frac{1}{2}S^0,$$

and thus, $-0.47 \leq E(Y_{a=1}) - E(Y_{a=0}) \leq -0.005$. We could then conclude from this study that antihistamine use truly had a beneficial effect on asthma. To draw conclusions about bounds for the causal effect using signed edges, [VanderWeele et al. \(2008\)](#) had to assume that the $W \rightarrow U$, $W \rightarrow A$, $U \rightarrow A$ and $U \rightarrow Y$ edges all had positive sign. From these assumptions, they concluded that $E(Y_{a=1}) - E(Y_{a=0}) \leq S^1 - S^0$ and from this it follows that $E(Y_{a=1}) - E(Y_{a=0}) \leq 0.06 - 0.25 = -0.19$. The application of Theorem 2 required assumptions about bounds related to only one edge, namely the $U \rightarrow Y$ edge.

As is clear from Example 2, Theorem 2 in this paper allows the researcher to draw conclusions about bounds on causal effects in the presence of unmeasured confounding by making assumptions concerning fewer edges than were previously required. [VanderWeele et al. \(2008\)](#) showed that to draw conclusions about the sign of the bias in the presence of unmeasured confounding using signed edges the treatment had to be binary or comparison had to be made between the minimum and maximum levels of treatment. Intuition about the sign of the bias could fail if an intermediate level of treatment was considered. In contrast, Theorem 2 makes no assumptions on whether A is binary, ordinal or continuous.

The approach employed in Example 2 to address bounds under unmeasured confounding by use of Theorem 2 is broadly applicable. In cases in which the outcome Y is binary, as in many epidemiologic studies, the approach is straightforward because specifying (Λ_L, Λ_H) that satisfy (2) consists only of specifying bounds on the risk ratio for the effect of U on Y and the inequalities in (3) of Theorem 2 then give bounds on the causal effect. By specifying bounds on the risk ratio for the effect of U on Y one immediately obtains bounds for the effect of A on Y .

ACKNOWLEDGEMENT

The authors thank the editor and two referees for helpful comments. VanderWeele acknowledges support from the National Institutes of Health, U.S.A. Tan acknowledges support from the National Science Foundation, U.S.A.

APPENDIX

[VanderWeele & Robins \(2009\)](#) proved Lemmas A1–A3 below, which will be used in this Appendix. Lemmas A2 and A3 are given in a somewhat more general form in [VanderWeele & Robins \(2009\)](#) but these special cases will suffice for our purposes here.

LEMMA A1. *If $h(z_2, z_1, q)$ is nondecreasing in z_1 and in z_2 and $\text{pr}(Z_2 > z_2 | Z_1 = z_1, Q = q)$ is nondecreasing in z_1 , for all z_2 , then $E\{h(Z_2, z_1, q) | Z_1 = z_1, Q = q\}$ is nondecreasing in z_1 .*

LEMMA A2. *Let X denote some set of nondescendants of Z_1 that block all back-door paths from Z_1 to Z_2 . If all directed paths between Z_1 and Z_2 are of positive sign then $\text{pr}(Z_2 > z_2 | z_1, x)$ is nondecreasing in z_1 for all z_2 and all x .*

LEMMA A3. *Suppose that Z_1 is a nondescendant of Z_2 and let Q denote the union of (i) the ancestors of Z_1 and (ii) the ancestors of Z_2 which are not descendants of Z_1 . Let $V_0 = Z_1$ and $V_n = Z_2$ and let*

$(V_1, \dots, V_{n-1})$ be an ordered list of all the nodes on directed paths from Z_1 to Z_2 exclusive of Z_1 and Z_2 then $\text{pr}(V_k > v_k \mid z_1, \bar{v}_{k-1}, q) = \text{pr}(V_k > v_k \mid \text{pa}_{v_k})$ for $k = 1, \dots, n$.

Proof of Proposition 1. Note that $0 \leq \Lambda_i^- < \infty$ and $0 < \Lambda_i^+ \leq \infty$ ($i = 1, \dots, n$). We give a proof for the inequality involving Λ_i^- . The inequality involving Λ_i^+ can be proved similarly by exchanging the roles of a_0 and a_1 . Recall that $\bar{V}_k$ denotes $(V_1, \dots, V_k)$.

We first establish the inequality for $n = 1$. Let $G^{-1}(g) = \sup\{v_1 : G(v_1) \leq g\}$. A useful result is that (i) if $G\{G^{-1}(g)\} \leq g$, then $G(v_1) > g$ if and only if $v_1 > G^{-1}(g)$ and that (ii) if $G\{G^{-1}(g)\} > g$, then $G(v_1) > g$ if and only if $v_1 \geq G^{-1}(g)$. We prove (i) by contradiction. Suppose $v_1 > G^{-1}(g)$. If $G(v_1) \leq g$, then $v_1 \leq G^{-1}(g)$ by the definition of $G^{-1}(g)$, which is a contradiction. Suppose $G(v_1) > g$. If $v_1 \leq G^{-1}(g)$, then $G(v_1) \leq G\{G^{-1}(g)\} \leq g$ since G is nondecreasing in v_1 , again a contradiction. To prove (ii), if $v_1 > G^{-1}(g)$, then $G(v_1) > g$, because if $G(v_1) \leq g$, then $v_1 \leq G^{-1}(g)$ by the definition of $G^{-1}(g)$, which is a contradiction, and if $v_1 = G^{-1}(g)$ then $G(v_1) > g$. To show the converse, suppose $G(v_1) > g$. If $v_1 < G^{-1}(g)$, then there exists u_1 such that $v_1 < u_1 < G^{-1}(g)$ and $G(u_1) \leq g$ by the definition of $G^{-1}(g)$ and this contradicts that G is nondecreasing. Due to (i) and (ii), we have that $\{v_1 : G(v_1) > g\}$ equals either $\{v_1 : v_1 > G^{-1}(g)\}$ or $\{v_1 : v_1 \geq G^{-1}(g)\}$.

Consider the case $\{v_1 : G(v_1) > g\} = \{v_1 : v_1 > G^{-1}(g)\}$. Then $\text{pr}\{G(V_1) > g \mid A = a_1, Q = q\} = \text{pr}\{V_1 > G^{-1}(g) \mid A = a_1, Q = q\}$ and $\text{pr}\{G(V_1) > g \mid A = a_0, Q = q\} = \text{pr}\{V_1 > G^{-1}(g) \mid A = a_0, Q = q\}$. The desired inequality is trivial by condition (a). Next, consider the case $\{v_1 : G(v_1) > g\} = \{v_1 : v_1 \geq G^{-1}(g)\}$. Then $\text{pr}\{G(V_1) > g \mid A = a_1, Q = q\} = \text{pr}\{V_1 \geq G^{-1}(g) \mid A = a_1, Q = q\}$ and $\text{pr}\{G(V_1) > g \mid A = a_0, Q = q\} = \text{pr}\{V_1 \geq G^{-1}(g) \mid A = a_0, Q = q\}$. Note that $p_1/p_0 \geq \Lambda_1^-$ if and only if $p_1 \geq \Lambda_1^- p_0$, for $0 \leq \Lambda_1^- < \infty$ and $0 \leq p_0, p_1 \leq 1$. Let v_{1j} be a sequence increasing to $G^{-1}(g)$. Then $\text{pr}(V_1 > v_{1j} \mid A = a_1, Q = q) \geq \Lambda_1^- \text{pr}(V_1 > v_{1j} \mid A = a_0, Q = q)$ by condition (a). Let $j \rightarrow \infty$, we obtain $\text{pr}\{V_1 \geq G^{-1}(g) \mid A = a_1, Q = q\} \geq \Lambda_1^- \text{pr}\{V_1 \geq G^{-1}(g) \mid A = a_0, Q = q\}$ and thus $\text{pr}\{G(V_1) > g \mid A = a_1, Q = q\} / \text{pr}\{G(V_1) > g \mid A = a_0, Q = q\} \geq \Lambda_1^-$.

Suppose that inequality (1) involving Λ_i^- holds for $n = k$. We show that it then holds for $n = k + 1$. Let $G_{\bar{v}_k}^{-1}(g) = \sup\{v_{k+1} : G(v_{k+1}, \bar{v}_k) \leq g\}$. Similarly as in the case for $n = 1$, we have that $\{v_{k+1} : G(v_{k+1}, \bar{v}_k) > g\}$ equals either $\{v_{k+1} : v_{k+1} > G_{\bar{v}_k}^{-1}(g)\}$ or $\{v_{k+1} : v_{k+1} \geq G_{\bar{v}_k}^{-1}(g)\}$ and, by condition (a), $\text{pr}\{G(V_{k+1}, \bar{v}_k) > g \mid \bar{V}_k = \bar{v}_k, A = a_1, Q = q\} \geq \Lambda_{k+1}^- \text{pr}\{G(V_{k+1}, \bar{v}_k) > g \mid \bar{V}_k = \bar{v}_k, A = a_0, Q = q\}$. Therefore,

$$\begin{aligned} \text{pr}\{G(\bar{V}_{k+1}) > g \mid A = a_1, Q = q\} &= E[\text{pr}\{G(V_{k+1}, \bar{V}_k) > g \mid \bar{V}_k, A = a_1, Q = q\} \mid A = a_1, Q = q] \\ &\geq \Lambda_{k+1}^- E[\text{pr}\{G(V_{k+1}, \bar{V}_k) > g \mid \bar{V}_k, A = a_0, Q = q\} \mid A = a_1, Q = q] \\ &= \Lambda_{k+1}^- \int_0^1 \text{pr}\{\Psi_g(\bar{V}_k) > z \mid A = a_1, Q = q\} dz, \end{aligned}$$

where $\Psi_g(\bar{v}_k) = \text{pr}\{G(V_{k+1}, \bar{v}_k) > g \mid \bar{V}_k = \bar{v}_k, A = a_0, Q = q\}$. Since $G(v_{k+1}, \bar{v}_k)$ is nondecreasing in v_{k+1} and $\bar{v}_k$ and since $\text{pr}(V_{k+1} > v_{k+1} \mid \bar{V}_k = \bar{v}_k, A = a_0, Q = q)$ is nondecreasing in $\bar{v}_k$, we have by Lemma A1 that $\Psi_g(\bar{v}_k)$ is nondecreasing in $\bar{v}_k$. Since (1) holds for $n = k$ we have that

$$\begin{aligned} \text{pr}\{G(\bar{V}_{k+1}) > g \mid A = a_1, Q = q\} &= \Lambda_{k+1}^- \int_0^1 \text{pr}\{\Psi_g(\bar{V}_k) > z \mid A = a_1, Q = q\} dz \\ &\geq \Lambda_{k+1}^- \int_0^1 \left(\prod_{i=1}^k \Lambda_i^- \right) \text{pr}\{\Psi_g(\bar{V}_k) > z \mid A = a_0, Q = q\} dz \\ &= \left(\prod_{i=1}^{k+1} \Lambda_i^- \right) \text{pr}\{G(\bar{V}_{k+1}) > g \mid A = a_0, Q = q\}, \end{aligned}$$

and so (1) holds for $n = k + 1$. □

Proof of Theorem 1. Fix $a_1 > a_0$. Since $\Theta(A, V_i) = (\Lambda_i^-, \Lambda_i^+)$ we have

$$\Lambda_i^- \leq \frac{\text{pr}(V_i > v_i \mid A = a_1, \text{pa}_Y^A)}{\text{pr}(V_i > v_i \mid A = a_0, \text{pa}_Y^A)} \leq \Lambda_i^+,$$

for all v_i and pa_Y^A , where pa_Y^A denote the parents of V_i other than A . Let Q denote the union of (i) the ancestors of A and (ii) the ancestors of Y which are not descendants of A then by Lemma A3, $\text{pr}(V_i > v_i \mid a, \text{pa}_Y^A) = \text{pr}(V_i > v_i \mid \bar{v}_{i-1}, a, q)$ and thus

$$\Lambda_i^- \leq \frac{\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_1, Q = q)}{\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a_0, Q = q)} \leq \Lambda_i^+ \quad (i = 1, \dots, n+1),$$

for all $v_i, \bar{v}_{i-1}, q$. Under the assumption that for all i such that there is an edge from A to V_i all directed paths from V_i to Y are of positive sign, relevant nodes can be replaced by their negations so that for all i such that there is an edge from A to V_i all edges on all directed paths from V_i to Y are of positive sign. By Lemma A2, $\text{pr}(V_i > v_i \mid \bar{V}_{i-1} = \bar{v}_{i-1}, A = a, Q = q)$ is nondecreasing in $\bar{v}_{i-1}$ for all v_i, a, q since (A, Q) will block all back-door paths from $\bar{V}_{i-1}$ to V_i . By Proposition 1 we thus have that

$$\prod_{i=1}^{n+1} \Lambda_i^- \leq \frac{\text{pr}(G(\bar{v}_{n+1}) > g \mid A = a_1, Q = q)}{\text{pr}(G(\bar{v}_{n+1}) > g \mid A = a_0, Q = q)} \leq \prod_{i=1}^{n+1} \Lambda_i^+, \quad (\text{A1})$$

for all g and q . Furthermore,

$$\begin{aligned} \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, X = x\} &= E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, X = x, Q\} \mid A = a, X = x] \\ &= E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, Q\} \mid A = a, X = x] \\ &= E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, W\} \mid A = a, X = x], \end{aligned}$$

where W is the subset of Q which are parents of one of the nodes in $\bar{V}_{n+1}$. There can be no unblocked front-door paths from A to W given X since the nodes in W are not descendants of A and thus any front-door path from A to W will be blocked given X by a collider. All back-door paths from A to W are blocked given X since X blocks all back-door paths from A to Y . From this it follows that all paths from A to W are blocked given X and so W is conditionally independent of A given X and so we have

$$\begin{aligned} E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, W\} \mid A = a, X = x] &= E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, W\} \mid X = x] \\ &= E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, Q\} \mid X = x]. \end{aligned}$$

We have thus shown that

$$\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, X = x\} = E[\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a, Q\} \mid X = x]. \quad (\text{A2})$$

By (A1) we have that

$$\left(\prod_{i=1}^{n+1} \Lambda_i^- \right) \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_0, Q = q\} \leq \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_1, Q = q\} \quad (\text{A3})$$

and

$$\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_1, Q = q\} \leq \left(\prod_{i=1}^n \Lambda_i^+ \right) \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_0, Q = q\} \quad (\text{A4})$$

for all g and q . Taking conditional expectations given $X = x$ of (A3) and (A4) and making use of relation (A2) we have that

$$\left(\prod_{i=1}^{n+1} \Lambda_i^- \right) \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_0, X = x\} \leq \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_1, X = x\}$$

and

$$\text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_1, X = x\} \leq \left(\prod_{i=1}^n \Lambda_i^+ \right) \text{pr}\{G(\bar{v}_{n+1}) > g \mid A = a_0, X = x\},$$

for all g and x . This completes the proof. $\square$

Proof of Theorem 2. Let $S^a = \sum_c E(Y \mid A = a, C = c) \text{pr}(C = c)$. We then have that

$$\begin{aligned} S^a &= \int_c E(Y \mid A = a, C = c) dF(c) \\ &= \int_c \left\{ \int_u E(Y \mid A = a, C = c, U = u) dF(u \mid A = a, C = c) \right\} dF(c) \\ &\leq \int_c \left\{ \int_u \Lambda_H E(Y \mid A = a, C = c, U = u') dF(u \mid A = a, C = c) \right\} dF(c) \\ &= \Lambda_H \int_c E(Y \mid A = a, C = c, U = u') dF(c) \\ &= \Lambda_H \int_c \left\{ \int_u E(Y \mid A = a, C = c, U = u') dF(u \mid C = c) \right\} dF(c) \\ &\leq \Lambda_H \int_c \left\{ \int_u \frac{1}{\Lambda_L} E(Y \mid A = a, C = c, U = u) dF(u \mid C = c) \right\} dF(c) \\ &= \frac{\Lambda_H}{\Lambda_L} \int_{c,u} E(Y \mid A = a, C = c, U = u) dF(u, c) \\ &= \frac{\Lambda_H}{\Lambda_L} E(Y_a). \end{aligned}$$

Thus, $(\Lambda_L / \Lambda_H) S^a \leq E(Y_a)$. The proof that $E(Y_a) \leq (\Lambda_H / \Lambda_L) S^a$ is similar. $\square$

Proof of Proposition 2. We have that $E(Y \mid u, l) = \int_0^\infty \text{pr}(Y > y \mid u, l) dy \leq \int_0^\infty \Lambda \text{pr}(Y > y \mid u', l) dy = \Lambda E(Y \mid u', l)$. $\square$

REFERENCES

- BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley: New York.
- CORNFIELD, J., HAENSZEL, W., HAMMOND, E. C., LILIENTHAL, A. M., SHIMKIN, M. B. & WYNDER, L. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Nat. Cancer Inst.* **22**, 173–203.
- DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *Int. Statist. Rev.* **70**, 161–89.
- GEIGER, D., VERMA, T. S. & PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20**, 507–34.
- GREENLAND, S., PEARL, J. & ROBINS, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37–48.
- HAARVELMO, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11**, 1–12.
- LAURITZEN, S. (1996). *Graphical Models*. Oxford: Oxford University Press.
- LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. & LEIMER, H. G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491–505.
- LIN, D. Y., PSATY, B. M. & KRONMAL, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounding in observational studies. *Biometrics* **54**, 948–63.
- LIU, C.-L. & WELLMAN, M. P. (1998). Using qualitative relationships for bounding probability distributions. In *Proc. 14th Conf. Uncertainty Artif. Intel.*, pp. 346–53. Madison: Morgan Kaufmann.
- LIU, C.-L. & WELLMAN, M. P. (2004). Bounding probabilistic relationships in Bayesian networks using qualitative influences: methods and applications. *Int. J. Approx. Reason.* **36**, 31–73.
- PARSONS, S. (1995). Refining reasoning in qualitative probabilistic networks. In *Proc. 11th Conf. Uncertainty Artif. Intel.*, pp. 427–34. San Francisco: Morgan Kaufman.
- PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669–88.
- PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

- RENOOIJ, S. & VAN DER GAAG, L. C. (2008). Enhanced qualitative probabilistic networks for resolving trade-offs. *Artif. Intel.* **172**, 1470–94.
- ROBINS, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In *Highly Structured Stochastic Systems*, Eds. P. Green, N. L. Hjort and S. Richardson, pp. 70–81. New York: Oxford University Press.
- SPIRITES, P., GLYMOUR, C. & SCHEINES, R. (1993). *Causation, Prediction and Search*. New York: Springer.
- VANDERWEELE, T. J., HERNÁN, M. A. & ROBINS, J. M. (2008). Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* **19**, 720–8.
- VANDERWEELE, T. J. & ROBINS, J. M. (2009). Properties of monotonic effects on directed acyclic graphs. *J. Mach. Learn. Res.* **10**, 699–718.
- VANDERWEELE, T. J. & ROBINS, J. M. (2010). Signed directed acyclic graphs for causal inference. *J. R. Statist. Soc. B* **72**, 111–27.
- VERMA, T. & PEARL, J. (1988). Causal networks: semantics and expressiveness. In *Proc. 4th Workshop on Uncertainty in Artif. Intel.*, pp. 352–9. Reprinted in *Uncertainty in Artificial Intelligence*, Vol. 4, Eds. R. Shachter, T. S. Levitt and L. N. Kanal, pp. 69–76. Amsterdam: Elsevier.
- WELLMAN, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artif. Intel.* **44**, 257–303.
- WRIGHT, S. (1921). Correlation and causation. *J. Agric. Res.* **20**, 557–85.

[Received April 2009. Revised August 2011]