American Journal of Epidemiology
Vol. 166, No. 9
DOI: 10.1093/aje/kwm179
Advance Access publication August 16, 2007

Practice of Epidemiology

Directed Acyclic Graphs, Sufficient Causes, and the Properties of Conditioning on a Common Effect

Tyler J. VanderWeele[1] and James M. Robins[2]

[1] Department of Health Studies, Division of Biological Sciences, University of Chicago, Chicago, IL.
[2] Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA.

Received for publication October 27, 2006; accepted for publication May 22, 2007.

In this paper, the authors incorporate sufficient-component causes into the directed acyclic graph (DAG) causal framework in order to make apparent several properties of conditioning on a common effect. By incorporating sufficient causes on a graph, it is possible to detect conditional independencies within strata of the conditioning variable which are not evident on DAGs without the representation of sufficient causes. It is also possible to determine the sign of the conditional covariance of two causes when conditioning on their common effect if some knowledge of the sufficient-cause mechanisms for the common effect is available. The incorporation of sufficient causes within the DAG framework also allows for the representation of interactions on DAGs and for the unification of several different causal frameworks. For illustration, the results are applied to an example concerning the familial coaggregation of two disorders.

causality; collider stratification; directed acyclic graphs; epidemiologic methods; independence; interaction; sufficient causes

Abbreviations: DAG; directed acyclic graph; SCC; sufficient-component cause.

Directed acyclic graphs (DAGs) have been used in epidemiology to represent causal relations among variables, and they have been used extensively to determine which variables it is necessary to condition on in order to control for confounding (1–4). In some cases, conditioning on a common effect can introduce bias even when none was present without conditioning (2). Within the DAG framework, this is referred to as ''collider stratification bias.'' Greenland (5) argues that in many cases, collider stratification will induce quantitatively less bias than will traditional confounding. Beyond this, relatively little is understood about the consequences of conditioning on a common effect.

In this paper, we demonstrate how Rothman's sufficient-component cause (SCC) model (6) can be represented on a causal DAG and how doing so elucidates the properties of conditioning on a common effect. We first review the SCC and DAG frameworks. We then provide a motivating example concerning familial coaggregation. Theory is then developed concerning the incorporation of sufficient causes on DAGs and on conditional independence and conditional covariance properties of conditioning on a common effect. Several previous papers have focused on how the SCC framework is related to the potential-outcomes causal framework (7–12). Here our focus will be on the relation between the SCC framework and the DAG framework. We furthermore discuss how the theory developed helps unify these various causal frameworks. Finally, we return to the motivating example and show how the methods developed in this paper can be applied. We then conclude with some additional discussion.

Correspondence to Dr. Tyler J. VanderWeele, Department of Health Studies, University of Chicago, 5841 South Maryland Avenue, MC 2007, Chicago, IL 60637 (e-mail: vanderweele@uchicago.edu).

OVERVIEW OF CAUSAL FRAMEWORKS AND MOTIVATING EXAMPLE

Sufficient-component causes

Rothman's SCC framework conceptualized causation as a series of different causal mechanisms, each sufficient to bring about the outcome (6). These causal mechanisms Rothman called "sufficient causes." He conceived of them as minimal sets of actions, events, or states of nature which together initiated a process resulting in the outcome. For a particular outcome there would likely be many different sufficient causes, that is, many different causal mechanisms by which the outcome could come about. Each sufficient cause involved various component causes. Whenever all components of a particular sufficient cause were present, the outcome would inevitably occur; within every sufficient cause, each component would be necessary for that sufficient cause to lead to the outcome.

We will use the following notation. An event is a binary variable taking values in $\{0, 1\}$. The odds ratio operator, $\vee$, is defined for two events A and B such that $A \vee B = 1$ if and only if either $A = 1$ or $B = 1$. The complement of an event A will be denoted by $\overline{A}$. A conjunction or product of the events $X_1, \ldots, X_n$ will be written as $X_1 \ldots X_n$, so that $X_1 \ldots X_n = 1$ if and only if each of the events $X_1, \ldots, X_n$ takes the value 1. Under the SCC framework (6), a series of events or conditions or causes, $F_1, \ldots, F_m$, all of which are binary, is said to be a *sufficient cause* for D if the series $F_1 \ldots F_m = 1$ implies that $D = 1$. If $S_1, \ldots, S_n$ are all of the sufficient causes for D, where each S_i is made up of some product of components which are binary, $S_i = F_1^i \ldots F_{m_i}^i$, so that $D = S_1 \vee \ldots \vee S_n$, then we will say that $S_1, \ldots, S_n$ are *determinative* for D.

Directed acyclic graphs

A DAG is composed of variables (nodes) and arrows between nodes (directed edges) such that the graph is acyclic—that is, 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* DAG is one in which the arrows can be interpreted as causal relations 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 DAGs can be found in the work of Greenland et al. (2). Greater formalization is provided in Pearl's work (1, 13), in which DAGs are considered graphical representations of structural equations such that each variable is defined as a function of its parents and a random error term.

Statistical associations on causal DAGs 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. We will graphically represent conditioning by placing a box around the variable on the graph upon which we are conditioning.

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 which follows the edges in the direction indicated by the graph's arrows. A collider is a node on a particular path such that both the preceding and subsequent nodes on the path have directed edges going into that node—that is, both the edge to and the edge from that node have arrowheads into the node. Note that a collider is relative to a particular path: A node that is a collider on one path may not be a collider on another path. A path between A and B is said to be blocked given (i.e., conditioning on) some set of variables Z if either there is a variable in Z on the path that is not a collider or 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 (14–16).

Motivating example

Consider a study in which each observation consists of two persons within the same family for whom data are available regarding two diseases: bipolar disorder, denoted by P, and binge eating, denoted by B. Suppose further that the two diseases are such that P could cause B but B could not cause P. For example, it is possible that bipolar disorder may lead to binge eating but rather implausible that binge eating would lead to bipolar disorder. The presence of bipolar disorder in persons 1 and 2 is denoted by P_1 and P_2, respectively. Similarly, the presence of binge eating in persons 1 and 2 is denoted by B_1 and B_2, respectively. Let E_i denote a certain factor particular to individual i. Let G_P denote some factor common to the family which is a cause of bipolar disorder but not of binge eating; let G_B denote some factor common to the family which is a cause of binge eating but not of bipolar disorder; and let F denote some set of factors common to the family which are causes of both bipolar disorder and binge eating. These causal relations are summarized in the causal DAG given in figure 1. The presence of factors F is said to constitute familial coaggregation (17, 18). Further detail concerning this specific example is given elsewhere (James Hudson et al., Harvard University, unpublished manuscript).

Suppose that data are available only on P_1, P_2, B_1, and B_2 and we wish to test the null hypothesis of no familial coaggregation (i.e., the null hypothesis that there are no directed edges emanating from F). Suppose further that E_1 and E_2 are never preventive for P_1 and B_1 or for P_2 and B_2, respectively, and that G_P is never preventive for P_1 or P_2 and that G_B is never preventive for B_1 or B_2. Later in this paper, we will show how tests for the null hypothesis of no familial coaggregation can be derived from the theory developed herein.

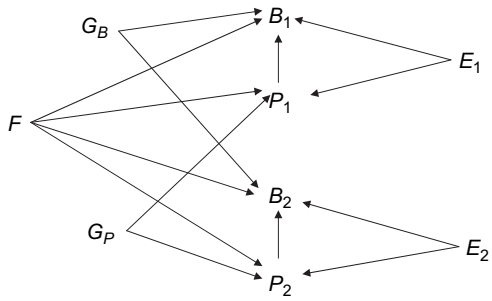


FIGURE 1. A causal directed acyclic graph with familial coaggregation.

SUFFICIENT CAUSATION STRUCTURES ON DAGS

If some node D is such that D and all of its parents are binary and the sufficient causes for D are known, it is possible to construct a new causal DAG with the sufficient causes for D on the DAG. Consider the causal DAG given in figure 2.

Suppose that the node D and all of its parents, $E_1, \ldots, E_5$, are binary. Suppose further that E_1E_2 and $\overline{E_2}E_3E_4$ and $E_4\overline{E_5}$ are a determinative set of sufficient causes for D. Then it can be shown (see Result 1 below) that the diagram given in figure 3 with all of the sufficient causes for D as new nodes is also a causal DAG. To indicate that the set of sufficient causes is determinative, we will add to the diagram an ellipse around the sufficient-cause nodes.

The example we have given is legitimized by and is a special case of the result presented below. The proof of this result can be found elsewhere (19).

Result 1. Consider a causal DAG G with some node D, such that D and all of its parents are binary. If a determinative set of sufficient causes for D, say $S_1, \ldots, S_n$, can be constructed from the parents of D and their complements, then a new causal DAG J can be formed by adding to G the nodes $S_1, \ldots, S_n$, removing the directed edges into D from the parents of D on G, adding directed edges from each S_i into D, and adding directed edges into each S_i from every parent of D on G which appears in the conjunction for S_i.

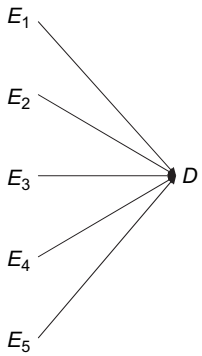


FIGURE 2. A causal directed acyclic graph without a sufficient causation structure.

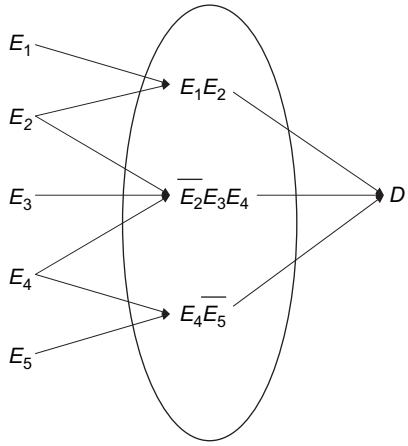


FIGURE 3. A causal directed acyclic graph with a sufficient causation structure.

When we construct these new causal DAGs with the sufficient causes, we will generally replace the sufficient-cause nodes S_i with the conjunctions that constitute them. We will call the resulting diagram a causal DAG with a *sufficient causation structure*. We will say that the node D *admits a sufficient causation structure*. One criticism of the causal DAG framework is that it does not allow for the representation of interactions among variables (9). However, causal DAGs with sufficient causation structures overcome this criticism by allowing for the graphical representation of interactions on a DAG. For example, in figure 3, E_1 and E_2 interact synergistically in their effects on D because they are both present in a single sufficient cause. The nodes E_2 and E_3 interact antagonistically in their effects on D, since E_2 and E_3 are both present in a single sufficient cause. Thus, in the case of binary variables, causal DAGs with sufficient causation structures overcome a major shortcoming in the traditional DAG causal framework.

If it is not possible to form a determinative set of sufficient causes for D from the parents of D and their complements, it may be possible to add nodes to the DAG so that on the larger causal DAG a determinative set of sufficient causes for D can be constructed from the parents of D and their complements. We will use the term *background causes* or *co-causes* to refer to the additional parents of D, say $A_0, \ldots, A_u$, which are added to the graph so as to be able to construct a determinative set of sufficient causes for D from the parents of D and their complements. Unless it is known that the set of parents $A_0, \ldots, A_u$ have no common causes, then a variable U with directed edges into each of $A_0, \ldots, A_u$ must also be added to the graph. It can be shown that it is always possible to find such additional nodes $A_0, \ldots, A_u$. In the Appendix, we show that it is always possible to find such additional nodes for two binary causes. The proof of the more general result can be found elsewhere (19).

The result given above provides a link between all four of the causal model frameworks discussed by Greenland and Brumback (9): graphical models, potential-outcome

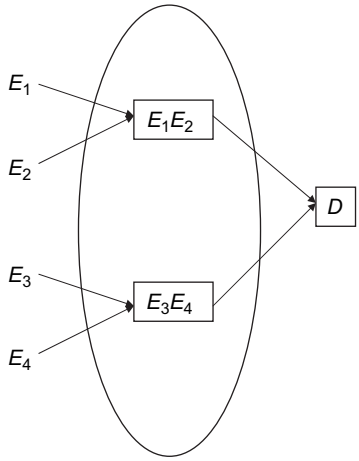


FIGURE 4. A sufficient causation structure with conditional independencies within the $D = 0$ stratum.

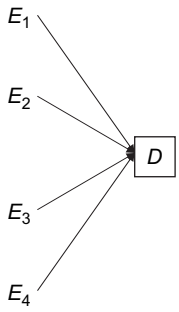


FIGURE 5. A causal directed acyclic graph with conditional independencies within the $D = 0$ stratum that are not evident from the *d*-separation criterion.

(counterfactual) models, SCC models, and structural-equation models. The four are linked through structural equations. Graphical models can be interpreted as diagrammatic shorthand for structural equations (1). Structural equations can be interpreted as sets of counterfactual relations (1, 7). The result presented above provides the final link by relating SCC models to graphical models and thereby also structural-equation models. In fact, the structural equation for each sufficient-cause node S_i is given by the product of components that constitute the sufficient cause $S_i = F^i_1 \ldots F^i_{m_i}$, since all of the components are needed for the sufficient cause to be realized. The structural equation for D is given by the disjunction of the sufficient-cause nodes $D = S_1 \vee \ \ldots \ \vee S_n$, since any of the sufficient causes suffices for the outcome D. Structural equations may thus be seen as a framework encompassing all four of these approaches to representing causal relations.

The construction of determinative sets of sufficient conjunctions for D will generally not be unique. For example, if $D = A_0 \vee A_1E$, it is also the case that $D = B_0 \vee B_1E$, where $B_0 = A_0$ and $B_1 = \overline{A_0}A_1$. This non-uniqueness of the sufficient causes for D is discussed further in the following section. If the parents of D on the original DAG are labeled $E_1, \ldots, E_m$, then each sufficient cause S_i must include either the variable E_i in its conjunction or $\overline{E_i}$ in its conjunction or must include neither E_i nor $\overline{E_i}$ in its conjunction; clearly it cannot include both. There are thus 3^m possible combinations of the E_i's and their complements that may appear in sufficient causes.

Conditional independence when conditioning on a common effect

Because a causal DAG with a sufficient causation structure is itself a causal DAG, the *d*-separation criterion applies and allows one to determine independencies and conditional independencies. A sufficient causation structure will often make apparent conditional independencies within strata of the conditioning variable which were not apparent on the original causal DAG. This is so because if some node D on a causal DAG admits a sufficient causation structure, then conditioning on $D = 0$ also conditions on all sufficient-cause nodes for D on the causal DAG with the sufficient causation structure. For example, consider a causal DAG with the sufficient causation structure given in figure 4.

Conditioning on $D = 0$ also conditions on $E_1E_2 = 0$ and $E_3E_4 = 0$, and thus we have by the *d*-separation criterion that, for example, E_1 is conditionally independent of E_4 given $D = 0$. This is because any path from E_1 to E_4 passes through E_1E_2, which is in the conditioning set, and therefore all paths between E_1 and E_4 are blocked given $D = 0$. If the causal DAG did not have the sufficient causation structure so that the causal relations were simply those given in figure 5, the conditional independence of E_1 and E_4 given $D = 0$ would no longer be apparent from the causal DAG. This is because the path from E_1 to E_4 through D is no longer blocked given $D = 0$, since we are conditioning on the collider D.

Our results provide the theoretical framework for and the generalization of the conditional independence example of Hernán et al. (20). To ensure that the DAG with the sufficient causation structure is itself a *causal* DAG, it is important that the set of sufficient causes for D on the graph be a determinative set of sufficient causes—that is, that the sufficient causes represent all of the pathways by which the outcome D may occur. Otherwise certain nodes may have common causes which are not on the graph, and the graph will then not be a *causal* DAG. For instance, appendix figure 2 in the paper by Hernán et al. (20) presents an example in which the causal structures are those indicated in figure 6.

In this example, a surgical procedure E affects survival through the removal of a tumor, and haplotype U affects survival through increasing levels of low density lipoprotein cholesterol, resulting in an increased risk of heart attack (regardless of whether a tumor is present). There are two cause-specific mortality variables: death from a tumor D_{1A} and death from a heart attack D_{1B}, either of which is sufficient for death D. Hernán et al. (20) claim that if death by tumor and death by heart attack are independent in the sense that they do not share a common cause and if surgery E is

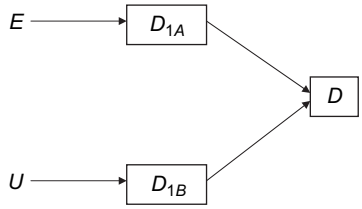


FIGURE 6. An example of the effects of surgery *E* and haplotype *U* on death *D*, from appendix figure 2 in the paper by Hernán et al. (20).

independent of haplotype U, then E and U will be conditionally independent given $D = 0$ (i.e., among the survivors). They make this claim on the basis of the DAG given in figure 6. The *d*-separation criterion would imply that E and U are conditionally independent given $D = 0$, since conditioning on $D = 0$ also conditions on $D_{1A} = 0$ and any path from E to U must pass through D_{1A}. However, even under the assumptions of Hernán et al. (20), the DAG in figure 6 may not be a *causal* DAG; this is because the nodes D and D_{1A} may have common causes even if E and U are independent and if D_{1A} and D_{1B} have no common causes; similarly, D and D_{1B} may have common causes. Consider, for example, the causal DAG with a sufficient causation structure given in figure 7, with $D_{1A} = A_1E$ and $D_{1B} = A_2U$.

Here A_0 represents the causes of death other than tumor and heart attack; A_1 represents the causes of death by tumor other than surgery E; A_2 represents the causes of heart attack other than haplotype U. In figure 6, D_{1A} and D_{1B} are marginally independent and have no common causes, the background causes A_1 and A_2 are marginally independent, and E and U are marginally independent. However, the background causes A_0 and A_1 have the common cause C_1 and the background causes A_0 and A_2 have the common cause C_2. Conditioning on $D = 0$ also conditions on $A_0 = 0, A_1E = 0$, and $A_2U = 0$, but, conditioning on $D = 0$, there is still an unblocked path from E to U, namely $E - D_{1A} - A_1 - C_1 - A_0 - C_2 - A_2 - D_{1B} - U$. The example thus illustrates a case not considered in the conditional independence example of Hernán et al. (20). It further demonstrates the importance of ensuring that the set of sufficient causes for D displayed on the graph is a determinative set of sufficient causes so that the resulting diagram is in fact a causal DAG—that is, so that all common causes of any two variables on the graph are also on the graph. If the sufficient causes which are added to the DAG are not a determinative set, the resulting diagram may not in fact be a causal DAG.

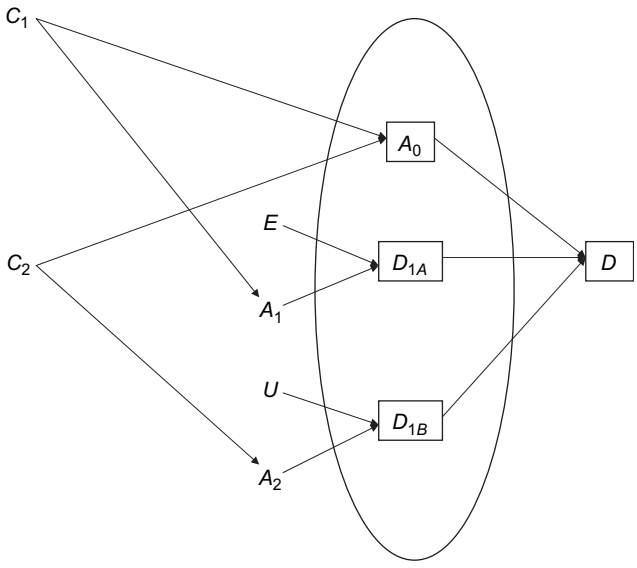


FIGURE 7. A sufficient causation structure in which *E* and *U* are independent and have independent co-causes but are not conditionally independent given $D = 0$.

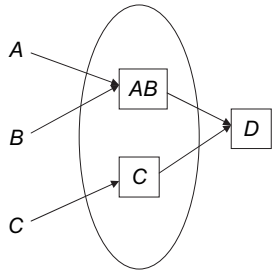


FIGURE 8. A sufficient causation structure with conditional independencies within the $D = 0$ stratum that are evident from the *d*-separation criterion.

We noted above that the set of determinative sufficient causes for D will not generally be unique. Consider the causal DAG with the sufficient causation structure indicated in figure 8.

Suppose that A, B, and C represent three toxic exposures such that A and B jointly or C alone is sufficient for the outcome D, death. Conditioning on $D = 0$ conditions also on $AB = 0$ and $C = 0$, and by the *d*-separation criterion, A is conditionally independent of C given $D = 0$. Suppose that the causal mechanisms are as follows: The presence of A and B jointly always causes heart failure resulting in death; and in the absence of B, the toxic exposure C always causes respiratory failure resulting in death; but in the presence of B, C causes a failure of the nervous system, again resulting in death. We then have three distinct causal mechanisms for death: AB, BC, and $\overline{B}C$. This implies that we can represent these causal mechanisms by means of the causal DAG with sufficient causation structure given in figure 9.

Both figures 8 and 9 are causal DAGs that correctly describe the causal relations among the variables; they differ in the level of detail present. In figure 9, conditioning on $D = 0$ also conditions on $AB = 0$, $BC = 0$, and $\overline{B}C = 0$, but the *d*-separation criterion no longer implies that A and C are conditionally independent given $D = 0$, because on the causal DAG there are two unblocked paths between A and C conditioning on $D = 0$, namely $A - AB - B - BC - C$ and $A - AB - B - \overline{B}C - C$. Thus, from the causal DAG given in figure 8, it was possible to use the *d*-separation criterion to

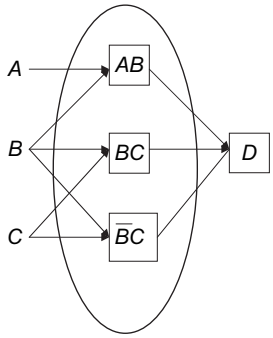


FIGURE 9. A sufficient causation structure with conditional independencies within the $D = 0$ stratum that are not evident from the *d*-separation criterion.

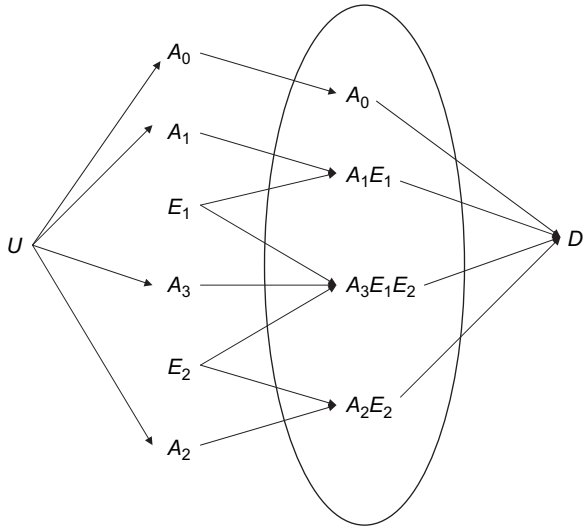


FIGURE 10. A sufficient causation structure in which E_1 and E_2 are independent and have positive monotonic effects on D.

identify the conditional independence of A and C given $D = 0$. However, from the causal DAG given in figure 9, the *d*-separation criterion would not identify this conditional independence relation, even though the two DAGs describe the same causal structure and even though the conditional independence relation truly does hold. The difficulty arises because on the graph in figure 9 we are representing sufficient causes which are not minimally sufficient. The mechanisms BC and $\overline{B}C$ are distinct—death due to failure of the nervous system versus death due to respiratory failure—but in neither mechanism is B or $\overline{B}$ necessary; C itself is always sufficient for death, regardless of whether the toxic exposure B is present. We see then that allowing sufficient causes which are not minimally sufficient on a causal DAG can sometimes obscure conditional independence relations. It may thus be desirable to use the causal DAGs given in both figures 8 and 9: the first to make clear the conditional independence relations and the second to represent the distinct causal mechanisms, which (interestingly) need not be minimally sufficient.

Conditional covariance when conditioning on a common effect

If some knowledge of the sufficient causes for an outcome D is available, it will sometimes be possible to determine the sign of the conditional covariance of two causes when conditioning on a common effect. First, however, we must introduce the concept of a *monotonic effect*. Consider an outcome D with two causes of interest, E_1 and E_2. We will say that E_1 has a *positive monotonic effect* on D if intervening to increase E_1 will never decrease D for any person, regardless of the level to which E_2 is set. We can define a monotonic effect for E_2 on D similarly. *Negative monotonic effects* are also defined analogously. Thus, the definition of a monotonic effect essentially requires that the effect of some intervention be in a particular direction for every person in the population, not merely on average. 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, one will be able to attribute a positive monotonic effect; whenever the intervention is always harmful or neutral for all individuals, one will be able to attribute a negative monotonic effect.

We now consider the relation of a monotonic effect to the sufficient causes of D. When D and all of its parents are binary, the presence of a monotonic effect of E on D implies that there exists a determinative set of sufficient causes for D such that $\overline{E}$ never appears in the conjunctions for any of the sufficient causes (19). Thus, if two parents of D, say E_1 and E_2, both have monotonic effects on D, then there exists a determinative set of sufficient causes for D such that the sufficient causes for D may have among their components E_1 or E_2 or E_1E_2 or neither E_1 nor E_2, but never $\overline{E_1}$ or $\overline{E_2}$ or $\overline{E_1}\overline{E_2}$. If E_1 and E_2 are independent of one another and have positive monotonic effects on D, the sufficient causation structure for the causal DAG can be described by the graph given in figure 10.

With this definition of a monotonic effect, we can now give Result 2.

Result 2. Suppose that D and all of its parents are binary, and suppose that two parents of D, say E_1 and E_2, are independent and also independent of all other parents of D on the causal DAG. Suppose further that E_1 and E_2 have a positive monotonic effect on D. Then, for any determinative set of sufficient causes D such that $D = A_0 \vee A_1E_1 \vee A_2E_2 \vee A_3E_1E_2$, the following hold:

1. If $A_0 = 0$, then $\mathrm{Cov}(E_1, E_2|D) \leq 0$.
2. If $A_0 = 0$ and A_1 and A_2 are independent, then $\mathrm{Cov}(E_1, E_2|\overline{D}) \leq 0$.
3. If $A_1 = 1$ or $A_2 = 1$, then $\mathrm{Cov}(E_1, E_2|D) \leq 0$ and $\mathrm{Cov}(E_1, E_2|\overline{D}) = 0$.
4. If $A_1 = 0$ or $A_2 = 0$, then $\mathrm{Cov}(E_1, E_2|D) \geq 0$ and $\mathrm{Cov}(E_1, E_2|\overline{D}) \leq 0$.

5. If $A_3 = 0$, then $\text{Cov}(E_1, E_2|D) \leq 0$.
6. If $A_3 = 0$ and if A_1 and A_2 are independent and additionally A_0 is independent of either A_1 or A_2, then $\text{Cov}(E_1, E_2|\overline{D}) = 0$.

The assumption that $A_0 = 0$ is essentially the assumption that one of E_1 or E_2 is always necessary for the outcome D; that is, D cannot occur without either E_1 or E_2. The assumption that $A_1 = 1$ is the assumption that E_1 is itself always sufficient for D; similarly, the assumption that $A_2 = 1$ is the assumption that E_2 is itself always sufficient for D. The assumption that $A_1 = 0$ is the assumption that E_1 by itself is never necessary for D; that is, if $E_1 = 1$ and $D = 1$, it must be the case either that $E_2 = 1$ or that $D = 1$, even if $E_1 = 0$. Similarly, the assumption that $A_2 = 0$ is the assumption that E_2 by itself is never necessary for D; that is, if $E_2 = 1$ and $D = 1$, it must be the case either that $E_1 = 1$ or that $D = 1$, even if $E_2 = 0$. The assumption that $A_3 = 0$ is essentially that there is no synergism in the SCC sense between E_1 and E_2; that is, if $D = 1$ and $E_1 = E_2 = 1$, then either $E_1 = 1$ alone or $E_2 = 1$ alone would be sufficient for D. If any of these assumptions can be made, we can draw conclusions about the conditional covariance between E_1 and E_2. Result 2 can be generalized if E_1 and E_2 are not independent. The statement of this generalization and its proof are provided in supplementary material posted on the *Journal*'s website (http://www.aje.oxfordjournals.org). The result can also be generalized when the conditional covariance of two nodes which are not parents of D are considered (19). We give an example of one such generalization, Result 3, in the Appendix.

MOTIVATING EXAMPLE REVISITED

We now return to the motivating example introduced at the beginning of this paper and show how Results 2 and 3 can be used to derive a statistical test for the presence of no familial coaggregation. We will graphically represent the monotonic-effects relations indicated earlier by signs on the appropriate edges. The null hypothesis of no familial coaggregation can then be represented by the signed causal DAG given in figure 11 with no arrows out of F.

Under the null hypothesis of no familial coaggregation, by Result 2, $\text{Cov}(E_1, G_P|P_1 = 1) \leq 0$ if E_1 and G_P do not exhibit synergism. By Result 3 (see Appendix), we have $\text{sign}(\text{Cov}(B_1, P_2|P_1 = 1)) = \text{sign}(\text{Cov}(E_1, G_P|P_1 = 1))$. Thus, under the null hypothesis of no familial coaggregation, $\text{sign}(\text{Cov}(B_1, P_2|P_1 = 1)) = \text{sign}(\text{Cov}(E_1, G_P|P_1 = 1)) \leq 0$ if there is no synergism between E_1 and G_P in the SCC sense. Consequently, a test of the null $\text{Cov}(B_1, P_2|P_1 = 1) \leq 0$ is a joint test of no familial coaggregation and no synergism between E_1 and G_P. Similarly, a test of the null $\text{Cov}(B_2, P_1|P_2 = 1) \leq 0$ is a joint test of no familial coaggregation and no synergism between E_2 and G_P.

DISCUSSION

The primary contributions of this paper are a number of theoretical advances in representing and reasoning about causal relations. Specifically, we have provided a link between and have helped unify several different causal models: SCC, counterfactual, graphical, and structural-equation models. Doing so allowed us to derive a number of properties of conditioning on a common effect. We have shown that representing sufficient causes on a DAG can allow for the detection of conditional independence relations within strata of the conditioning variable that are not evident on traditional causal DAGs. We have also stated conditions which allow a researcher to draw conclusions about the sign of the conditional covariance of two causes when conditioning on a common effect. Finally, we have shown how sufficient causes and synergistic interactions can be graphically represented on causal DAGs.

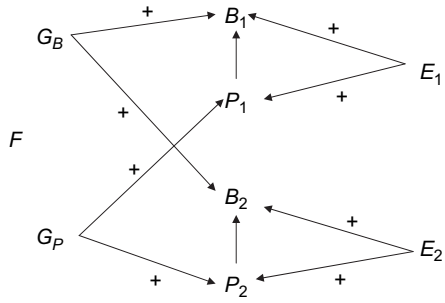


FIGURE 11. A causal directed acyclic graph with monotonic effects, under the null hypothesis of no familial coaggregation.

For the theory presented in this paper to be applied to epidemiologic problems, relatively strong assumptions are needed: binary outcomes and exposures, the conditional independence assumptions of DAGs, knowledge of or conjectures about determinative sets of sufficient causes, and monotonic effects. These assumptions limit the applicability of the theory. We will discuss each of these assumptions in turn.

Although the theory is limited to binary outcomes and exposures, many outcomes and exposures of interest in epidemiologic research are binary. Furthermore, for ordinal or categorical exposures, the theory developed can be applied by recoding the exposure as a series of binary variables; however, the outcome would still need to be binary in order for Results 2 and 3 to be applied. Our results made use of the various conditional independence assumptions entailed by DAGs. Many medical and epidemiologic systems can be usefully represented by DAGs, as testified to by the increasing occurrence of DAGs in the epidemiologic literature. The structure of the DAG itself implies the conditional independence assumptions which we made use of in our results. The requirement that the researcher have knowledge of or conjectures about a determinative set of sufficient causes can be relaxed somewhat, but the details are beyond the scope of the current paper (19). Monotonic effects assume that a particular exposure affects all persons the same way (i.e., the effect points in the same direction for everyone); although this is a strong assumption, it is one that may apply to a number of epidemiologic exposures, such as the effect

of smoking on lung cancer risk or the effect of certain genes or environmental exposures. Furthermore, the assumption of monotonic effects can sometimes be weakened to assumptions about the ordering of intervention distributions, as discussed in other work (21).

Although the assumptions required to apply the theory in this paper are considerable, they are not insurmountable, as our example from psychiatric epidemiology demonstrates. Moreover, our results perform an important cautionary function: They define the number, type, and strength of the assumptions that are required in order to succeed in using epidemiologic data to draw conclusions from sufficient causes on causal DAGs.

The DAG framework has proven to be a useful tool for causal thinking in epidemiologic research. Several recent papers have extended the applicability of DAGs to new types of problems (20–24). The contributions in this paper concerning sufficient causes and the properties of conditioning on a common effect extend further the scope of the types of problems which DAGs can address.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82:669–88.
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;12:313–20.
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;155:176–84.
5. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003;14:300–6.
6. Rothman KJ. Causes. Am J Epidemiol 1976;104:587–92.
7. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. Scand J Work Environ Health 1988;14:125–9.
8. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia, PA: Lippincott-Raven, 1998.
9. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol 2002;31:1030–7.
10. Flanders D. Sufficient-component cause and potential outcome models. Eur J Epidemiol 2006;21:847–53.
11. VanderWeele TJ, Hernán MA. From counterfactuals to SCCs, and vice versa. Eur J Epidemiol 2006;21:855–8.
12. VanderWeele TJ, Robins JM. The identification of synergism in the SCC framework. Epidemiology 2007;18:329–39.
13. Pearl J. Causality: models, reasoning, and inference. Cambridge, United Kingdom: Cambridge University Press, 2000.
14. Verma T, Pearl J. Causal networks: semantics and expressiveness. In: Shachter RD, Levitt TS, Kanal LN, et al. Uncertainty in artificial intelligence, 4. (Machine Intelligence and Pattern Recognition series, vol 9). Amsterdam, the Netherlands: Elsevier/North-Holland, 1990:69–76.
15. Geiger D, Verma TS, Pearl J. Identifying independence in Bayesian networks. Networks 1990;20:507–34.
16. Lauritzen SL, Dawid AP, Larsen BN, et al. Independence properties of directed Markov fields. Networks 1990;20:491–505.
17. Hudson JI, Laird NM, Betensky RA. Multivariate logistic regression for familial aggregation of two disorders. I. Development of models and methods. Am J Epidemiol 2001;153:500–5.
18. Hudson JI, Laird NM, Betensky RA, et al. Multivariate logistic regression for familial aggregation of two disorders. II. Analysis of studies of eating and mood disorders. Am J Epidemiol 2001;153:506–14.
19. VanderWeele TJ, Robins JM. Minimal sufficient causation and directed acyclic graphs. In: VanderWeele TJ. Contributions to the theory of causal directed acyclic graphs. (PhD thesis). Cambridge, MA: Harvard University, 2006:1–42.
20. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615–25.
21. 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, 2006:43–120.
22. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology 2006;17:360–72.
23. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight "paradox" uncovered? Am J Epidemiol 2006;164:1115–20.
24. VanderWeele TJ, Robins JM. Four types of effect modification—a classification based on directed acyclic graphs. Epidemiology 2007 (in press).

APPENDIX

Construction of Co-Causes for Two Binary Causes and a Binary Outcome

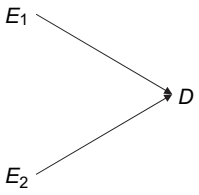
Suppose that E_1 and E_2 are the only parents of D on the original causal directed acyclic graph (DAG), as in appendix figure 1.

We will show that it is possible to construct co-causes which can be added to the DAG so that a determinative set of sufficient causes can be formed from $E_1, E_2, \overline{E_1}, \overline{E_2}$ and the co-causes. The construction we give will work even if E_1 and E_2 have common causes. We construct $A_0, \ldots, A_8$ so that $D = A_0 \vee A_1E_1 \vee A_2\overline{E_1} \vee A_3E_2 \vee A_4\overline{E_2} \vee A_5E_1E_2 \vee A_6\overline{E_1}E_2 \vee A_7E_1\overline{E_2} \vee A_8\overline{E}_1\overline{E}_2$ by defining the variables A_0, A_1, A_2, A_3, A_4, A_5, A_6, A_7, and A_8 as follows. Let $D_{ij}(\omega)$ be the counterfactual value of D for individual ω if E_1 were set to i and E_2 were set to j. Then:

let $A_0(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 0$ if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$;

let $A_1(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 3$ if $D_{00}(\omega) = 0$, $D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$;

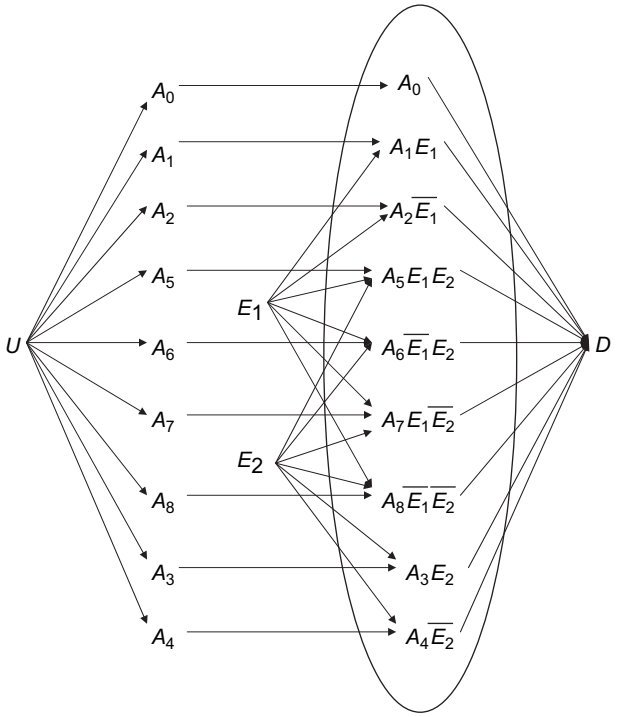
let $A_2(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2, 3$ if $D_{10}(\omega) = 0$, $D_{00}(\omega) = D_{01}(\omega) = D_{11}(\omega) = 1$;



APPENDIX FIGURE 1. A causal directed acyclic graph with two causes and no sufficient causation structure.

let $A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 3$ if $D_{00}(\omega) = D_{10}(\omega) = 0$, $D_{01}(\omega) = D_{11}(\omega) = 1$;
let $A_1(\omega) = A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 4$ if $D_{01}(\omega) = 0$, $D_{00}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$;
let $A_1(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1$ if $D_{00}(\omega) = D_{01}(\omega) = 0$, $D_{10}(\omega) = D_{11}(\omega) = 1$;
let $A_5(\omega) = A_8(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 5, 8$ if $D_{01}(\omega) = D_{10}(\omega) = 0$, $D_{00}(\omega) = D_{11}(\omega) = 1$;
let $A_5(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 5$ if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 0$, $D_{11}(\omega) = 1$;
let $A_2(\omega) = A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2, 4$ if $D_{11}(\omega) = 0$, $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 1$;
let $A_6(\omega) = A_7(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 6, 7$ if $D_{00}(\omega) = D_{11}(\omega) = 0$, $D_{01}(\omega) = D_{10}(\omega) = 1$;
let $A_2(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2$ if $D_{10}(\omega) = D_{11}(\omega) = 0$, $D_{00}(\omega) = D_{01}(\omega) = 1$;
let $A_6(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 6$ if $D_{00}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0$, $D_{01}(\omega) = 1$;
let $A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 4$ if $D_{01}(\omega) = D_{11}(\omega) = 0$, $D_{00}(\omega) = D_{10}(\omega) = 1$;
let $A_7(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 7$ if $D_{00}(\omega) = D_{01}(\omega) = D_{11}(\omega) = 0$, $D_{10}(\omega) = 1$;
let $A_8(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 8$ if $D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0$, $D_{00}(\omega) = 1$; and
let $A_i(\omega) = 0$ for all i if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0$.

The causal DAG with this sufficient causation structure is then that given in appendix figure 2.



APPENDIX FIGURE 2. A sufficient causation structure for D with two binary causes, E_1 and E_2.

Result 3. Suppose that E_1, E_2, and D are binary variables, that E_1 and E_2 are the only parents of D, that F and G are d-separated given $\{E_1, E_2, D\}$, that F and E_2 are d-separated given $\{E_1, D\}$, and that G and E_1 are d-separated given $\{E_2, D\}$. Suppose also that E_1 is a parent of F, that E_1 has a monotonic effect on F, that there are no intermediate variables between E_1 and F, and that E_1 and F have no common causes; and suppose that E_2 is a parent of G, that E_2 has a monotonic effect on G, that there are no intermediate variables between E_2 and G, and that E_2 and G have no common causes. Then $\text{sign}(\text{Cov}(F, G|D)) = \text{sign}(\text{Cov}(E_1, E_2|D))$.