

Four types of effect modification - a classification based on directed acyclic graphs

BY TYLER J. VANDERWEELE

Department of Health Studies, University of Chicago
5841 South Maryland Avenue, MC 2007, Chicago, IL 60637
vanderweele@uchicago.edu

AND JAMES M. ROBINS

Departments of Epidemiology and Biostatistics, Harvard School of Public Health
677 Huntington Avenue, Boston, MA 02115
robins@hsph.harvard.edu

Abstract. By expressing the conditional causal risk difference as a sum of products of stratum specific risk differences and conditional probabilities, it is possible to give a classification of the types of causal relationships that can give rise to effect modification on the risk difference scale. Directed acyclic graphs make clear the necessary causal relationships for a particular variable to serve as an effect modifier for the causal risk difference concerning two other variables. The directed acyclic graph causal framework thereby gives rise to a four-fold classification for effect modification: direct effect modification, indirect effect modification, effect modification by proxy and effect modification by a common cause. Brief discussion is given to the case of multiple effect modification relationships and multiple effect modifiers as well as measures of effect other than that of the causal risk difference.

Key Words: Causal inference; directed acyclic graphs; effect modification; interaction.

Directed acyclic graphs have been used as causal diagrams in epidemiologic research for a variety of purposes. Directed acyclic graphs have been used to represent causal relations amongst variables;¹⁻³ they have been used extensively to determine the variables for which it is necessary to control for confounding in order to estimate causal effects;^{1-2,4-5} more recently they have been used by Hernán et al. to provide a classification of the types of causal relationships that can give rise to selection bias.⁶ In this paper we follow their work by using directed acyclic graphs to provide a classification of the types of causal relationships that can give rise to effect modification. Specifically, we consider what relationships an effect modifier variable may exhibit in relation to the variable constituting the cause and the variable constituting the effect. Doing so yields a structural classification of effect modification; the classification is *structural* in that it makes reference to the structure of the causal directed acyclic graph. We first provide some discussion of the various measures of effect used to assess effect modification. We will then focus on the causal risk difference as a measure of effect by which effect modification is assessed (though much of the discussion applies also to other measures of effect as well) and we use directed acyclic graphs to provide a classification of different types of effect modification. Extensions to conditioning on multiple items and to scales other than the risk difference are discussed at the paper's conclusion and in Appendix 1.

Effect Modification

Epidemiologists apply the term "effect modification" to indicate that the effect of one variable on another varies across strata of a third. There are many different measures of effect and thus also many different measures by which a variable may be an effect modifier for the relationship between a cause and an effect. There has been considerable discussion as to which measure of effect one might most naturally use in assessing effect modification. The risk difference, the risk ratio and the odds ratio are all frequently used in assessing effect modification. In general, different measures of effect will be useful in different contexts. For example, the risk difference which, within the context of effect modification, measures departures from the additivity of effects is arguably of greatest public health importance;⁷⁻⁹ while the odds ratio is the natural measure of choice for case-control studies.

Here we will focus primarily on the causal risk difference as our measure of choice. By using the causal risk difference, the various relationships an effect modifier may bear to the

variable constituting the cause and the variable constituting the effect are made particularly clear. Although we focus on the causal risk difference much of the discussion applies also to other measures of effect. It should also be noted that effect modification on the causal risk difference scale or any other scale falls under the broader idea of "interaction." The relationship between effect modification and different notions of interaction in a counterfactual framework has been developed elsewhere.^{3,10–12} These issues are outside the intended scope of this paper. Much of this literature concerns individual-level interaction. Directed acyclic graphs allow for the graphical representation of population level causal relationships and thus the causal risk difference (or alternatively causal risk ratio or odds ratio) provides the most appropriate focus for our analysis.

Before proceeding, one further issue merits discussion. It is often commented that conditioning on intermediate variables between the exposure or intervention variable and the outcome variable will in general bias estimates of the causal effect.^{13–17} In considering which variables may act as effect modifiers, we will therefore restrict our attention to variables that are not a consequence of the exposure or intervention variable under consideration. A variable Q will thus be said to be an effect modifier on the causal risk difference scale of the relationship between some exposure E and some outcome D if Q is not affected by E and there exists two levels of E , e_0 and e_1 say, such that the difference between the expected value of D over the population intervening to set the exposure variable E to e_1 as compared with intervening to set the exposure variable E to e_0 varies across strata of Q .

More formally, let $D_{E=e}$ denote the counterfactual variable D intervening to set the exposure variable E , possibly contrary to fact, to level e . Then the causal effect of E on D comparing two levels of E , e_0 and e_1 , is defined simply as the causal risk difference $\mathbb{E}[D_{E=e_1}] - \mathbb{E}[D_{E=e_0}]$. And we thus say that a variable Q is an effect modifier for the causal risk difference of E on D if Q is not affected by E and if there exists two levels of E , e_0 and e_1 , such that $\mathbb{E}[D_{E=e_1}|Q = q] - \mathbb{E}[D_{E=e_0}|Q = q]$ is not constant in q . With this definition in place, we may proceed to classify different types of effect modification by using causal directed acyclic graphs.

Causal Directed Acyclic Graphs

A directed acyclic graph is composed of variables (nodes) and arrows between nodes (directed edges) 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. Thus if a graph has nodes V_1, \dots, V_n and some variable U is a common cause of say V_1 and V_2 then U must be included on the graph for the graph to be a causal directed acyclic graph. If, on the other hand U were only a cause of V_1 or only a cause of V_2 but not both, then U could be included on the graph or it could be left out. Its inclusion wouldn't be necessary for the graph to be a causal directed acyclic graph. That the graph's edges are directed ensures that causes precede effects; that the graph is acyclic ensures that no variable can be its own cause. If there is a directed edge from A to B then A is said to be a parent of B and B is said to be a child of A . If there are a series of one or more directed edges such that it is possible to begin at node A , follow the directed edges in the arrowhead direction and end at another node B then A is said to be an ancestor of B and B is said to be a

descendent of A . If A is a parent of B then A is a direct cause of B ; if A is an ancestor of B but not a parent of B then A is said to be an indirect cause of B (through the intermediate variables between A and B). Additional details can be found in the work of Greenland et al.² Greater formalization is provided by Pearl.¹

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 either a direct or indirect cause of B or if B is a direct or indirect 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 the strata of K ; that is to say A and B will in general be statistically associated given K if K is a common effect of A and B . We will graphically represent conditioning by placing a box around the variable on the graph upon which we are conditioning.⁵

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 particular node on a path such that both the preceding and subsequent nodes on the path have directed edges going into that node i.e. 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 . It has been shown that if all paths between A and B are blocked given Z then A and B are conditionally independent given Z .^{18–20}

We need one further result in the development of the structural classification of effect modification below. The backdoor path adjustment theorem states that for intervention variable E and outcome D , if a set of variables Z such that no variable in Z is a descendent of E blocks all "back-door paths" from E to D (i.e. all paths with directed edges into E) then conditioning on Z suffices to control for confounding for the estimation of the causal effect of E on D and this causal effect is given by $\mathbb{E}[D_{E=e}] = \sum_z \mathbb{E}[D|E=e, Z=z]P(Z=z)$.¹ Note that this is a graphical variant of Theorem 4 of Rosenbaum and Rubin²¹ and Rubin's g-formula.^{22–23}

A Structural Classification of Effect Modification

In this section we use causal directed acyclic graphs to consider what relationships an effect modifier variable may exhibit in relation to the variable constituting the cause and the variable constituting the effect. Doing so yields a classification of the different types of effect modification. Before presenting the theory for the classification we will illustrate the classification with a simple hypothetical example. Suppose that in a randomized trial, some drug E for hypertension D has variable effect (on the causal risk difference scale) according to the presence of some genotype X . Since the trial is randomized, E and X have no common causes and since both E and X have an effect on D the causal relationships among these variables can be described as those on the causal directed acyclic graph given in Figure 1.

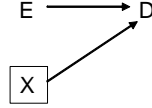


Figure 1. Direct Effect Modification: E - drug exposure; D - hypertension outcome; X - genotype, a direct effect modifier.

We could say that X is a direct effect modifier for the causal effect of E on D because X is a direct cause of D . Suppose now that information is available on the genotype C of the mothers of the study participants. The mothers' genotype affects the genotype of the study participants but does not affect hypertension of the study participants directly. The causal relationships among these variables is represented by the causal directed acyclic graph given in Figure 2.

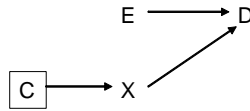


Figure 2. Indirect Effect Modification: E - drug exposure; D - hypertension outcome; X - genotype; C - mother's genotype, an indirect effect modifier.

We will show more formally below that in Figure 2, C will likely serve as an effect modifier on the causal risk difference scale for the effect of E on D . This is essentially because C affects X which serves as an effect modifier for the causal effect of E on D . We could thus say that C is an indirect effect modifier for the causal effect of E on D since C affects D indirectly through X . Now suppose that genotype X also determined skin pigment and that the study information on race R is based entirely on skin pigment. The causal relationships among these variables could then be represented by the causal directed acyclic graph given in Figure 3.

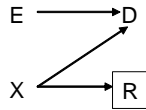


Figure 3. Effect Modification by Proxy: E - drug exposure; D - hypertension outcome; X - genotype; R - race, an effect modifier by proxy.

Here, R will also likely serve as an effect modifier on the causal risk difference scale for the effect of E on D because conditioning on R gives information on X which serves as an effect modifier for the causal effect of E on D . However, since R is not a cause of D we would say that R is an effect modifier of the causal effect of E on D by proxy. Finally, suppose that information is available on the race of the mother which we will denote by M . The causal relationships among the variables could then be represented by the causal directed acyclic graph given in Figure 4.

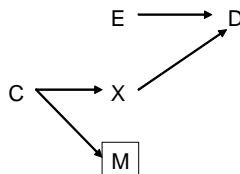


Figure 4. Effect Modification by Common Cause: E - drug exposure; D - hypertension outcome; X - genotype; C - mother's genotype; M - mother's race, an effect modifier by common cause.

It will be seen below that M also will likely serve as an effect modifier of the causal risk difference of E on D because conditioning on M gives information on C which affects X which serves as an effect modifier for the causal effect of E on D . Because C is a common cause of X (a variable that is a direct cause of D) and M (the variable we are conditioning on) we might refer to M as an effect modifier by common cause of the effect of E on D .

We now generalize this simple example and show that all instances of effect modification can be classified as falling into one of the four categories indicated above: direct effect modification, indirect effect modification, effect modification by proxy and effect modification by common cause. The classification is carried out by expressing the conditional causal risk difference as a sum of products of stratum specific risk differences and conditional probabilities as given in Theorem 1. In Theorem 1, we assume that there are no intermediate variables between E and D on the directed acyclic graph. In the next section, this assumption is dropped and the result is generalized. Proofs of all theorems are given in Appendix 2.

Theorem 1. Suppose that E is a parent of D and that there are no intermediate variables between E and D . Let X denote the parents of D other than E . Let Q be some set of non-descendants of E and D then

$$\begin{aligned} \mathbb{E}[D_{E=e_1}|Q=q] - \mathbb{E}[D_{E=e_0}|Q=q] \\ = \sum_x \{\mathbb{E}[D|X=x, E=e_1] - \mathbb{E}[D|X=x, E=e_0]\}P(X=x|Q=q). \end{aligned} \quad (1)$$

Theorem 1 states that the causal risk difference for D comparing two levels of E , e_1 and e_0 say, within a particular stratum of Q is given by the sum of the expected risk differences in D conditional on X and Q weighted by the probability of X given Q where X denotes the parents of D other than E . The formula given in (1) allows us to provide a structural classification of effect modification on the causal risk difference scale. For Q to be an effect modifier of the causal effect of E on D it is necessary that the function $G(q) = \mathbb{E}[D_{E=e_1}|Q=q] - \mathbb{E}[D_{E=e_0}|Q=q] = \sum_x \{\mathbb{E}[D|X=x, E=e_1] - \mathbb{E}[D|X=x, E=e_0]\}P(X=x|Q=q)$ is not constant in q . In other words, it is necessary that the expected risk difference in D conditional on X and $Q=q$ weighted by the probability of X given $Q=q$ is not constant in q . This latter expression depends on q only through $P(X=x|Q=q)$ and so it is necessary that $P(X=x|Q=q)$ is not constant in q . The requirement that $P(X=x|Q=q)$ is not constant in q is simply the requirement that X and Q are statistically associated. In the introductory material on directed acyclic graphs we discussed the various structures that may give rise to association between two variables: cause and effect, common causes, and conditioning on a common effect. We may thus use our knowledge of directed acyclic graphs to consider various cases for which a potential effect modifier Q will be associated with one or more of the variables in X . This will allow us to classify the type of effect modification for any potential effect modifier Q on the graph. Our analysis will follow the hypothetical example given above. First the conditioning variable may be among the variables in X (i.e. it may be a parent of D) and this gives rise to what we will call direct effect modification

as illustrated in Figure 1 with $Q = X$; second, the conditioning variable may be an ancestor of one or more of the variables in X which gives rise to what we will call indirect effect modification as illustrated in Figure 2 with $Q = C$; third, Q may be a descendent of one or more of the variables in X which gives rise to what we will call effect modification by proxy as illustrated in Figure 3 with $Q = R$; lastly, Q and one or more of the variables in X may have a common cause which gives rise to what we will call effect modification by a common cause as illustrated in Figure 4 with $Q = M$. Theorem 1 allowed us to transform the condition for effect modification on the causal risk difference scale into the necessary condition for effect modification that Q and X are statistically associated. Our knowledge of association structures on causal directed acyclic graphs then allowed us to classify different types of effect modification.

The four types of effect modification may be distinguished in a number of ways. First, as is clear from Figures 3 and 4 an effect modifier for the effect of some exposure on a particular outcome may not itself have a causal effect on that outcome. In the cases of direct and indirect effect modification, the effect modifier does have a causal effect on the outcome; in the cases of effect modification by proxy and by common cause the effect modifier does not. This is because the unblocked path from Q to X that gives rise to the required association between Q and X will be a frontdoor path from Q to X in the cases of direct or indirect effect modification and will be a backdoor path from Q to X in the cases of effect modification by proxy or by common cause. Second, direct effect modification may be distinguished from the other three types in an important way. If one is conditioning on multiple variables which include all the direct effect modifiers X then no other variable on the graph will continue to serve as an effect modifier for the causal effect of E on D while conditioning on X . This is essentially because X blocks all paths from any other potential effect modifier Q to D . In a sense, direct effect modifiers take precedence over all other types. The case of conditioning on multiple variables is considered further in the Discussion section and in Appendix 1.

One additional comment is necessary. For the function $G(q) = \sum_x \{\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]\}P(X = x|Q = q)$ to not be constant in q it is also necessary that the function $\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]$ is not constant in x . That is to say it is necessary that X be an effect modifier for the relationship between E and D . This will often, but not always, be the case. In the context of binary E and X , the expression $\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]$ will often not be constant in x if E and X exhibit synergism. Exceptions to the condition that $\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]$ is not constant in x will occur whenever all individual response types which Greenland and Poole classify as exhibiting "causal interdependence" are absent.¹⁰ Further discussion of some of these issues can be found in a companion paper and elsewhere.^{11,12,24} Here it suffices to note that if none of the variables in X serve as an effect modifier for the causal effect of E on D then no other variable on the graph will serve as an effect modifier because $\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]$ is then constant in x . Theorem 1 thus allows us to classify instances of effect modification but not to identify effect modification.

It is furthermore of note that it is possible for $P(X = x|Q = q)$ to not be constant in q and for $\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]$ to not be constant in x but to still have $G(q) = \sum_x \{\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]\}P(X = x|Q = q)$ constant in q (in which case Q would not be an effect modifier for the causal risk difference). This possibility arises because the differences $\{\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]\}$ may cancel

each other out perfectly. These will however be exceptional cases and in general whenever $P(X = x|Q = q)$ is not constant in q , effect modification on the risk difference scale will be manifest.

Theorem 1 made reference to a set of variables X constituted by the parents of D other than E . If E is the only parent of D then this set is empty and it turns out that there can then be no effect modifier on the graph for the causal effect of E on D . This is stated formally in Theorem 2.

Theorem 2. Suppose some node D on a causal directed acyclic graph has only one parent, E , then there exists no variable on the directed acyclic graph which is an effect modifier for the causal effect of E on D .

Essentially Theorem 2 states that if the exposure E is the only variable on the directed acyclic graph which is a direct cause of D then there can be no variable on the directed acyclic graph which acts as an effect modifier for the relationship between E and D . This is so because any other variable which could have an effect on D must do so through E but intervening on E will supersede any effect this other variable might otherwise have had. Note that the theorem does not state that there exists no effect modifier for the causal effect of E on D ; only that there is no variable on the directed acyclic graph which is an effect modifier for this relationship. A causal directed acyclic graph which included all the variables on the original causal directed acyclic graph plus a number of others might have a variable which could serve as an effect modifier for the causal effect of E on D if at least one of the additional variables were a direct cause of D (note that there may be causes of D which were not on the original causal directed acyclic graph if these variables are not also causes of another variable on the graph and thus not common causes of two or more variables on the graph). We see then that for there to be an effect modifier for the causal effect of E on D on a causal directed acyclic graph, the node D must have more than one parent.

Discussion

Several comments merit additional attention with regard to possible limitations and extensions of the results we have considered in this paper. First, it should be acknowledged that although effect modification is an important phenomenon, it is sometimes of insufficient magnitude to be of clinical or public health relevance. The theory developed in this paper does not allow for the representation of the magnitude of effect modification and thus cannot distinguish between cases in which the magnitude is or is not of substantive importance. Directed acyclic graphs are a useful conceptual tool but cannot take the place of empirical analysis and presentation of numerical results.

Second, effect modification relationships may be considerably more complicated than the examples given in Figures 1-4. The causal diagram may, for example, involve a number of intermediate variables. Furthermore, a variable in Q may be associated with more than one variable in X giving rise to multiple effect modification relationships. Additionally, the set Q itself may contain more than one variable thereby giving rise to multiple effect modifiers. These and other more complex causal structures and effect modification relationships are discussed in Appendix 1.

Third, although the analysis has been restricted to the use of the causal risk difference as a measure by which to assess effect modification, most of the remarks hold true if one considers other measures of effect. For example, the causal risk ratio for the causal effect of E on D comparing two levels of E , e_0 and e_1 , in stratum $Q = q$ is given by $\frac{\{\sum_x \mathbb{E}[D|X=x, E=e_1]P(X=x|Q=q)\}}{\{\sum_x \mathbb{E}[D|X=x, E=e_0]P(X=x|Q=q)\}}$. If X and Q are not independent then the expression $\frac{\{\sum_x \mathbb{E}[D|X=x, E=e_1]P(X=x|Q=q)\}}{\{\sum_x \mathbb{E}[D|X=x, E=e_0]P(X=x|Q=q)\}}$ will in most cases vary in q although exceptions can occur just as in the case of the causal risk difference above. Similarly, the causal odds ratio for the effect of E on D will generally vary in levels of Q if Q is not independent of X . The analysis however is most simple for the causal risk difference.

Fourth, for an exposure E , an outcome D and a potential effect modifier Q , several different causal directed acyclic graphs of varying complexity may represent the causal relationships among these variables. One graph with E , D and Q may have a number of additional variables; another may have only the variables E , D and Q and no others. The one requirement that must be satisfied with regard to the presence or absence of other variables on the graph is that any common cause of two variables on the graph must also be on the graph. Thus, for example, in Figure 2, the variable X could have been excluded from the graph with an arrow from C directly into D . The variable X could have been excluded because although it is a cause of D , it is not a cause of any other variable on the graph. In our system of classification, if X had been excluded from the graph then C would then be classified as a direct modifier of the causal effect of E on D . The classification of effect modifiers will thus sometimes be relative to the particular variables included on the graph under consideration. However, Figure 6 illustrates that indirect effect modification cannot always be reduced to direct effect modification by excluding variables from the graph. In Figure 6, Q_1 serves as an indirect effect modifier for the causal effect of E on D . However, the variable X_1 cannot be removed from the graph because it is a common cause of E and D . Similarly, in Figure 4, M was an effect modifier by common cause which could have been reduced to an effect modifier by proxy if X had been excluded from the graph with an arrow from C directly into D . However, in Figure 6, Q_2 serves as an effect modifier by common cause through X_2 but cannot be reduced to an effect modifier by proxy because X_2 cannot be removed from the graph as it is a common cause of E and D . How an effect modifier is classified is thus sometimes, but not always, relative to the particular variables that are represented on the causal directed acyclic graph under consideration.

Finally, we note that the results developed in this paper allowed us to classify any given instance of effect modification as one of direct effect modification, indirect effect modification, effect modification by proxy or effect modification by common cause. The theory did not pertain to identifying effect modification or to graphically representing interactions or effect modification or interactions rather merely to classifying instances of effect modification according to the structure of a directed acyclic graph. Theory concerning the graphical representation of synergistic interactions for binary variables is developed in related work.²⁵ The results in this paper classify and make clear the necessary causal relationships an effect modifier variable may and must exhibit in relation to the variable constituting the cause and the variable constituting the effect.

References

1. Pearl J. Casual diagrams for empirical research. *Biometrika* 1995;**82**:669-688.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiol.* 1999;**10**:37-48.
3. Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int. J. Epidemiol.* 2002;**31**:1030-1037.
4. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiol.* 2001;**12**:313-320.
5. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.* 2002;**155**:176-184.
6. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiol.* 2004;**15**:615-625.
7. Blot WJ, Day NE. Synergism and interaction: are they equivalent? *Am. J. Epidemiol.* 1979;**100**:99-100.
8. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am. J. Epidemiol.* 1980;**112**:467-470.
9. Saracci R. Interaction and synergism. *Am. J. Epidemiol.* 1980;**112**:465-466.
10. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand. J. Work Environ. Health* 1988;**14**:125-129.
11. Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven; 1998.
12. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component cause framework. *Epidemiol*, in press.
13. Greenland S, Neutra RR. Control of confounding in the assessment of medical technology. *Int. J. Epidemiol.* 1980;**9**:361-367.
14. Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J. Roy. Statist. Soc. Ser. A* 1984;**147**:656-666.
15. Robins JM, Morgenstern H. The foundations of confounding in epidemiology. *Math Modeling* 1987;**14**:869-916.
16. Robins JM. The control of confounding by intermediate variables. *Statist. Med.* 1989;**8**:679-701.

17. Weinberg CR. Toward a clearer definition of confounding. *Am. J. Epidemiol.* 1993;**127**:1-8.
18. Verma T, Pearl J. Causal networks: Semantics and expressiveness. In *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, 352-359. Reprinted in *Uncertainty in Artificial Intelligence* (ed. R. Shachter, T.S. Levitt, and L.N. Kanal), 1988;**4**:69-76. Amsterdam: Elsevier.
19. Geiger D, Verma TS, Pearl J. Identifying independence in bayesian networks. *Networks* 1990;**20**:507-534.
20. Lauritzen SL, Dawid AP, Larsen BN, Leimer HG. Independence properties of directed markov fields. *Networks* 1990;**20**:491-505.
21. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41-55.
22. Robins JM. A new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986;**7**:1393-1512.
23. 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. *Computers and Mathematics with Applications* 1987;**14**:923-945.
24. Koopman JS. Causal models and sources of interaction. *Am. J. Epidemiol.* 1977;**106**:439-444.
25. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes and the properties of conditioning on a common effect. *Submitted*.

Appendix 1: More Complex Effect Modification Structures.

As noted in the Discussion section, effect modification relationships may be considerably more complicated than the examples given in Figures 1-4. Any number of variables might lie between Q and X in Figure 2 or between X and Q in Figure 3 or between C and X or C and Q in Figure 4; the set X may contain more than one variable; and, as discussed below, there may be intermediate variables between E and D . However, when one considers the relation between the potential effect modifier Q and some particular member of the set X with which it is associated, their association will arise from one of the four alternatives presented above. Note however that these four alternatives are not mutually exclusive. A variable in Q may be associated with more than one variable in X and may exhibit any one of these four relations to (or be independent of) each particular variable in X . An example in which multiple effect modification relationships are present is given in Figure 5.

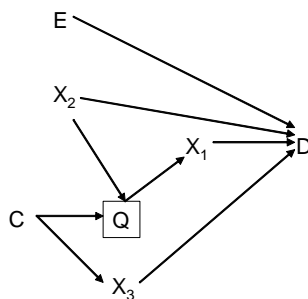


Figure 5. Multiple Effect Modification Relationships: E - exposure; D - outcome; X_1 , X_2 , X_3 - other direct causes of D ; Q - an effect modifier; C - a common cause.

In the rather complicated example given in this figure, Q is an effect modifier for the causal effect of E on D indirectly through X_1 , by proxy through X_2 , and by common cause through X_3 . The four-way classification above is sufficient if attention is restricted to conditioning on a single item. But it is also possible that a researcher is interested in two variables considered jointly as an effect modifier for the causal effect of E on D . One can then provide a classification of the effect modification structure for each variable in the set Q . An example of multiple variables serving as effect modifiers is given in Figure 6.

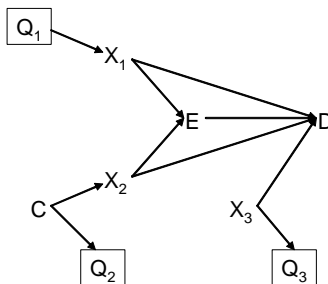


Figure 6. Multiple Effect Modifiers: E - exposure; D - outcome; X_1 , X_2 , X_3 - other direct causes of D ; Q_1 - an indirect effect modifier; Q_2 - an effect modifier by common cause; Q_3 - an effect modifier by proxy; C - a common cause.

In this example, the set X (the parents of D other than E) consists of the variables X_1 , X_2 and X_3 . The variable Q_1 serves as an indirect effect modifier through X_1 ; the variable Q_2

is an effect modifier by common cause through X_2 ; the variable Q_3 is an effect modifier by proxy through X_3 . The variables X_1 , X_2 and X_3 , if they had been conditioned upon (instead of Q_1 , Q_2 and Q_3) would all have served as direct effect modifiers of the causal effect of E on D . Of course, more complicated arrangements are also possible in which each of the variables in the set Q exhibits multiple effect modification relationships to the variables in X and in which additional intermediate variables are present.

We have seen that X and Q may be associated either because of a cause and effect relationship (Q directly causes X , Q indirectly causes X , or X causes Q) or through a common cause of X and Q ; this led to the fold-four classification of effect modification given above. However, as noted in the introductory material on directed acyclic graphs, X and Q may also be associated by conditioning on a common effect of these two groups of variables. This might occur in practice if, for example, one were interested in two variables considered jointly as an effect modifier. Alternatively this might occur if one were interested in only one variable as an effect modifier but, due to the sampling procedure, one intentionally or inadvertently conditioned on a particular subset of subjects which restricted the sample to a particular stratum of one of the causal directed acyclic graph's variables. In such cases, the common effect being conditioned upon, say K , might open a previously blocked path between Q and some variable in X . We might then still apply the four-fold classification given above with K now taking on the role of Q for purposes of classification. It turns out, however, that when Q is an effect modifier for the causal effect of E on D because of conditioning on a common effect of X and Q , this will rule out the cases of direct and indirect effect modification; effect modification will always be either by proxy or by common cause. If the effect modification were direct or indirect (rather than by proxy or common cause) then the unblocked path from K to X would be a frontdoor path and conditioning on K would then block the path from Q to X (unless Q were also associated with X in ways other than paths through K). Thus the relationship between K and X must either be that of proxy or of common cause. The relationship between Q and K may either be that Q is an ancestor of K or that Q and K share a common cause. In summary, when conditioning on K , we must thus have either effect modification by proxy (conditioning on a common effect, with Q and K related by ancestry or by a common cause) or effect modification by common cause (conditioning on a common effect, with Q and K related by ancestry or by a common cause). Examples of each of these four cases are given in the directed acyclic graphs presented in Figures 7-10.

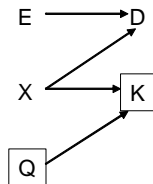


Figure 7. Effect Modification by Proxy (conditioning on a common effect with Q and K related by ancestry): E - exposure; D - outcome; X - a direct cause of D ; K - a conditioning variable; Q - an effect modifier by proxy.

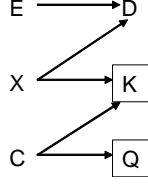


Figure 8. Effect Modification by Proxy (conditioning on a common effect with Q and K related by a common cause): E - exposure; D - outcome; X - a direct cause of D ; K - a conditioning variable; C - a common cause; Q - an effect modifier by proxy.

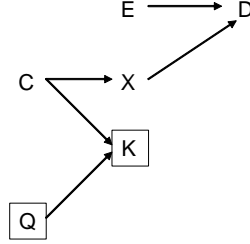


Figure 9. Effect Modification by Common Cause (conditioning on a common effect with Q and K related by ancestry): E - exposure; D - outcome; X - a direct cause of D ; K - a conditioning variable; C - a common cause; Q - an effect modifier by common cause.

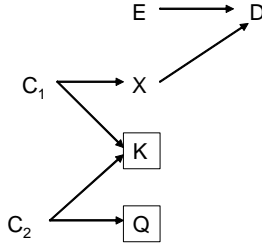


Figure 10. Effect Modification by Common Cause (conditioning on a common effect with Q and K related by a common cause): E - exposure; D - outcome; X - a direct cause of D ; K - a conditioning variable; C_1, C_2 - common causes; Q - an effect modifier by common cause.

Conditioning on two or more items along with the effect modifier may complicate matters yet further but the same principles apply. It may be that Q is associated with X only by conditioning on several other variables. Classification may take place by considering the first conditioning variable on a particular unblocked path between X and Q . Each consecutive pair of conditioning items or the final conditioning item and Q may be related by ancestry or by common cause. As was the case in Figures 5 and 6, multiple effect modification relationships or multiple effect modifiers might also be present.

Our examples and the discussion thus far have assumed that there are no intermediate variables between E and D on the directed acyclic graph. The results and discussion, however, easily generalize to the setting in which there are intermediate variables between E and D on the directed acyclic graph. Theorem 3 restates Theorem 1 dropping the assumption of no intermediate variables between E and D on the directed acyclic graph. The conclusion of Theorem 2 is the same as that of Theorem 1 but the conditions under which this conclusion holds are slightly different.

Theorem 3. Let D be some node on a causal directed acyclic graph with ancestor E and let X denote all non-descendents of E which are either parents of D or parents of a node on

a directed path between E and D . Let Q be some set of non-descendants of E and D then equation (1) holds.

In the presence of intermediate variables between E and D on the directed acyclic graph, effect modification can be classified as before according to their relationships with the set of variables X . However, in the context of intermediate variables between E and D on the directed acyclic graph, the set X is no longer simply the parents of D other than E but rather all non-descendants of E which are either parents of D or parents of a node on a directed path between E and D .

A final warning is necessary. Theorems 1 and 3 did not allow for Q to be a descendent of D . Consider the example given in Figure 11 in which Q is a descendent of D .

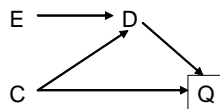


Figure 11. An example illustrating conditioning on a descendent of the outcome: E - exposure; D - outcome; X - a direct cause of D ; Q - a descendent of D .

When Q is a descendent of D then the causal effect of E on D conditioning on Q is no longer given by equation (1) because conditioning on Q provides information on D other than that which is available through X and E . In such cases Q does not serve as a genuine effect modifier for the causal effect of E on D because it is a consequence of D .

Appendix 2: Proofs.

Proof of Theorem 1.

Theorem 1 is a consequence of Theorem 3 below.

Proof of Theorem 2.

Let Q be some non-descendent of E and D then $\mathbb{E}[D_{E=e_1}|Q = q] - \mathbb{E}[D_{E=e_0}|Q = q] = \mathbb{E}[D_{E=e_1}] - \mathbb{E}[D_{E=e_0}]$ by Theorem 3 of Pearl¹ since $(D \coprod Q|E)_{G_{\overline{E}}}$ where $G_{\overline{E}}$ denotes the graph obtained by deleting from the original directed acyclic graph all arrows pointing into E . Furthermore, $\mathbb{E}[D_{E=e_1}] - \mathbb{E}[D_{E=e_0}] = \mathbb{E}[D = 1|E = 1] - \mathbb{E}[D = 1|E = 0]$ by the backdoor path adjustment theorem since there are no unblocked back-door paths from E to D as E is the only parent of D . Thus $\mathbb{E}[D_{E=e_1}|Q = q] - \mathbb{E}[D_{E=e_0}|Q = q] = \mathbb{E}[D = 1|E = 1] - \mathbb{E}[D = 1|E = 0]$ which is independent of q .

Proof of Theorem 3.

By the law of iterated expectations we have $\mathbb{E}[D_{E=e_1}|Q = q] - \mathbb{E}[D_{E=e_0}|Q = q] = \sum_x \mathbb{E}[D_{E=e_1}|X = x, Q = q]P(X = x|Q = q) - \sum_x \mathbb{E}[D_{E=e_0}|X = x, Q = q]P(X = x|Q = q)$. We will show that this latter expression is equal to $\sum_x \mathbb{E}[D_{E=e_1}|X = x]P(X = x|Q = q) - \sum_x \mathbb{E}[D_{E=e_0}|X = x]P(X = x|Q = q)$. By Theorem 3 of Pearl¹ it suffices to show that $(D \coprod Q|X, E)_{G_{\overline{E}}}$ where $G_{\overline{E}}$ denotes the graph obtained by deleting from the original directed acyclic graph all arrows pointing into E . Any front door path from D to Q in $G_{\overline{E}}$ will be blocked by a collider. Any backdoor path from D to Q in $G_{\overline{E}}$ will be blocked by X . We thus have that $\mathbb{E}[D_{E=e_1}|Q = q] - \mathbb{E}[D_{E=e_0}|Q = q] = \sum_x \mathbb{E}[D_{E=e_1}|X = x]P(X = x|Q = q) - \sum_x \mathbb{E}[D_{E=e_0}|X = x]P(X = x|Q = q)$. Since X will block all backdoor paths from E to D we have by the backdoor path adjustment theorem $\sum_x \mathbb{E}[D|X = x, E = e_1]P(X = x|Q = q) - \sum_x \mathbb{E}[D|X = x, E = e_0]P(X = x|Q = q) = \sum_x \{\mathbb{E}[D|X = x, E = e_1] - \mathbb{E}[D|X = x, E = e_0]\}P(X = x|Q = q)$.