Collection of Biostatistics Research Archive COBRA Preprint Series

Year 2006

Paper 12

Biologic Interaction and Their Identification

Tyler J. VanderWeele* James Robins[†]

*University of Chicago, tvanderw@hsph.harvard.edu

[†]Harvard School of Public Health, robins@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/cobra/art12

Copyright ©2006 by the authors.

Biologic Interaction and Their Identification

Tyler J. VanderWeele and James Robins

Abstract

The definitions of a biologic interaction and causal interdependence are reconsidered in light of a sufficient-component cause framework. Various conditions and statistical tests are derived for the presence of biologic interactions. The conditions derived are sufficient but not necessary for the presence of a biologic interaction. Through a series of examples it is made evident that in the context of monotonic effects, but not in general, the conditions which are derived are closely related but not identical to effect modification on the risk difference scale.

1. Introduction

The distinction between a biologic interaction (or synergism) and a statistical interaction has frequently been noted (Blot and Day, 1979; Rothman et al., 1980; Saracci, 1980). In the case of binary variables, concrete attempts have been made to articulate which types of counterfactual response patterns would constitute instances of interdependent effects (Miettinen, 1982; Miettinen, 1985; Greenland and Poole, 1988). In what follows we reconsider the definition of causal interdependence and its relation to that of biologic interaction or synergism in light of the sufficient-component cause framework (Rothman, 1976). Consideration of this framework gives rise to a definition of "definite interdependence" which constitutes a sufficient but not necessary condition for the presence of a biologic interaction. We then derive various statistical tests for the presence of biologic interactions and give a number of examples which illustrate the difference between the concepts of definite interdependence and effect modification on the risk difference scale.

2. The Definition of a Biologic Interaction

Suppose that D and two of its causes, E_1 and E_2 , are binary variables taking values 0 or 1. In the discussion that follows E_1 and E_2 are treated symmetrically so that E_1 could be relabeled as E_2 and E_2 could be relabeled as E_1 . We assume a deterministic counterfactual model. 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. For event E we will denote the complement of the event by \overline{E} . The probability of an event E occurring, P(E = 1), we will frequently simply denote by P(E). If there were some individual ω for whom $D_{10}(\omega) =$ $D_{01}(\omega) = D_{00}(\omega) = 0$ but for whom $D_{11}(\omega) = 1$ we would say that there was present a biologic interaction between the effect of E_1 and E_2 on D because in such a case there exists an individual for whom E_1 or E_2 alone is insufficient for D but for whom E_1 and E_2 together yield D. There is thus joint action between E_1 and E_2 and so we would speak of a biologic interaction. Similarly if there were individuals for whom $D_{11}(\omega) = D_{01}(\omega) = D_{00}(\omega) = 0$ and $D_{10}(\omega) = 1$; or for whom $D_{11}(\omega) = D_{10}(\omega) = D_{00}(\omega) = 0$ and $D_{01}(\omega) = 1$; or for whom $D_{11}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ we would again say that a biologic interaction was present. In the first of these three additional cases, there is a biologic interaction because only E_1 and $\overline{E_2}$ together imply D; in the second case because only $\overline{E_1}$ and E_2 together imply D; and in the third case because only $\overline{E_1}$ and $\overline{E_2}$ together imply D. We have considered four different response patterns which manifest what might be called a biologic interaction. We will see below that these four response patterns and in fact two others are closely related to synergism within the sufficient-component cause framework.

Miettinen (1982, 1985) classified the various response patterns which arise from two binary causes, E_1 and E_2 , and a binary outcome D into sixteen different response types according to the individuals' counterfactual outcomes as enumerated in Table 1.

COBRA A BEPRESS REPOSITORY Collection of Biostatistics Research Archive

Type	$E_1 = 1, E_2 = 1$	$E_1 = 0, E_2 = 1$	$E_1 = 1, E_2 = 0$	$E_1 = 0, E_2 = 0$
1	1	1	1	1
2	1	1	1	0
3	1	1	0	1
4	1	1	0	0
5	1	0	1	1
6	1	0	1	0
7	1	0	0	1
8	1	0	0	0
9	0	1	1	1
10	0	1	1	0
11	0	1	0	1
12	0	1	0	0
13	0	0	1	1
14	0	0	1	0
15	0	0	0	1
16	0	0	0	0

Table 1 Enumeration of response patterns to four possible exposure combinations.

Types 8, 10, 12, and 14 were classified by Miettinen as instances of causal interdependence. Types 3, 5, 7 and 9 were classified as instances of preventive interdependence. Miettinen thus included types 3, 5, 7, 8, 9, 10, 12, and 14 as those which constituted interdependent effects. Greenland and Poole (1988) criticized this classification because it was not invariant to interchanging the reference categories (i.e. relabeling for E_1 or for E_2 the label "1" as "0" and "0" as "1"). Type 15 for instance which is not classified as exhibiting causal interdependence in Miettinen's system would become a type 12 responder (which Miettinen did classify as exhibiting causal interdependence) if $E_2 = 0$ were relabeled $E_2 = 1$ and vice versa. Greenland and Poole therefore partitioned the types into equivalence classes which were invariant under the recoding of exposure indicators. Under the classification of Greenland and Poole, the equivalence class of types 7 and 10 is invariant and is said to exhibit mutual antagonism; the class composed of types 8, 12, 14 and 15 is invariant and consists of those types in which disease occurs for only one exposure combination; the class of types 2, 3, 5 and 9 is invariant and consists of those types in which disease occurs for exactly three exposure combinations. Under the classification of Greenland and Poole, these three classes constituting types 2, 3, 5, 7, 8, 9, 10, 12, 14 and 15 are all said to represent causal interdependence. Greenland and Poole note that if none of types 2, 3, 5, 7, 8, 9, 10, 12, 14 or 15 are present then the causal risk difference will be additive so that $\mathbb{E}[D_{11}] - \mathbb{E}[D_{00}] = (\mathbb{E}[D_{10}] - \mathbb{E}[D_{00}]) + (\mathbb{E}[D_{01}] - \mathbb{E}[D_{00}])$ where \mathbbm{E} denotes the average in the study population. It may be discerned, however, that Greenland and Poole's classification is insufficiently stringent for associating causal interdependence with a biologic interaction.

In the analysis that follows we will frequently use the disjunctive or OR operator, \bigvee , which is defined by $a \bigvee b = a + b - ab$ and is such that $a \bigvee b = 1$ if and only if either a = 1 or b = 1. A conjunction or product of the events $X_1, ..., X_n$ will be written as $X_1...X_n$ so that $X_1...X_n = 1$ if and only if each of the the events $X_1, ..., X_n$ takes the value 1. Under the sufficient-component cause framework (Rothman 1976), if $S_1, ..., S_n$ are all the sufficient causes for D then $D = S_1 \bigvee ... \bigvee S_n$ and each S_i is made up of some product of components, $F_1^i, ..., F_{m_i}^i$, which are binary so that $S_i = F_1^i ... F_{m_i}^i$. Following Rothman (1976; Koopman, 1981), we will say that two causes, E_1 and E_2 , for some outcome D, exhibit synergism (or a biologic interaction) if E_1 and E_2 are ever present together or "co-participate" in the same sufficient cause. If E_1 and $\overline{E_2}$ are present together in the same sufficient cause then the two causes again have a biologic interaction and E_1 and E_2 are said to exhibit antagonism; in this case it could also be said that E_1 and $\overline{E_2}$ exhibit synergism.

There are certain correspondences between response types and sets of sufficient causes. Green-

land and Poole (1988), in the case of two binary causes, enumerate nine different sufficient causes each involving some combination of E_1 and E_2 and their complements along with certain binary background causes. We may label these background causes as $A_0, A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8$. The nine different sufficient causes Greenland and Poole give are then $A_0, A_1E_1, A_2\overline{E_1}, A_3E_2, A_4\overline{E_2}, A_5E_1E_2, A_6\overline{E_1}E_2, A_7E_1\overline{E_2}$ and $A_8\overline{E_1}\overline{E_2}$. We thus have that

$$D = A_0 \bigvee A_1 E_1 \bigvee A_2 \overline{E_1} \bigvee A_3 E_2 \bigvee A_4 \overline{E_2} \lor A_5 E_1 E_2 \lor A_6 \overline{E_1} E_2 \lor A_7 E_1 \overline{E_2} \lor A_8 \overline{E_1} \overline{E_2}.$$

If one of A_5, A_6, A_7, A_8 were non-zero, it would then be proper to speak of a biologic interaction between E_1 and E_2 .

Knowing whether there is a biologic interaction between E_1 and E_2 will in general require having some knowledge of the causal mechanisms for the outcome D. For although a particular set of sufficient causes along with the presence or absence of the various background causes A_0 , A_1 , A_2 , $A_3, A_4, A_5, A_6, A_7, A_8$ for a particular individual suffices to fix a response type (Greenland and Poole 1988), the converse is not true (Greenland and Brumback 2002). That is to say, knowledge of an individual's response type does not generally fully determine which background causes are present. As an example, an individual who has $A_1(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 3$ has a sufficient cause completed if and only if $E_1 = 1$ (in which case A_1E_1 is completed) or $E_2 = 1$ (in which case A_3E_2 is completed). For such a individual we could write $D = E_1 \bigvee E_2$. Thus this individual would be of response type 2 because the individual will escape disease only if exposed to neither E_1 nor E_2 so that no sufficient cause is completed. In contrast, knowledge of a individual's response type does not generally fully determine which background causes are present. A individual who is of response type 2 could have either $A_1(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 3$ in which case we could write $D = E_1 \bigvee E_2$ or alternatively such a individual may have $A_5(\omega) = A_6(\omega) =$ $A_7(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 5, 6, 7$ in which case we could write $D = E_1 E_2 \bigvee \overline{E_1} E_2 \bigvee \overline{E_1} \overline{E_2}$. As noted by Greenland and Brumback (2002), it is thus impossible in this case to distinguish from the counterfactual response pattern alone the set of sufficient causes $E_1 \bigvee E_2$ from the set of sufficient causes $E_1 E_2 \bigvee \overline{E_1} E_2 \bigvee \overline{E_1} \overline{E_2}$. With both sets of sufficient causes, D will occur when either E_1 or E_2 is present. Whether $E_1 \bigvee E_2$ or $E_1 E_2 \bigvee \overline{E_1} E_2 \bigvee E_1 \overline{E_2}$ represent the proper description of the causal mechanisms for D can only be resolved with knowledge of the subject matter in question.

Using the sufficient cause representation for D given above we can see that Greenland and Poole's (1988) classification of those types which represent causal interdependence is insufficiently stringent for associating causal interdependence with a biologic interaction. Greenland and Poole include types 2, 3, 5 and 9 amongst those types that are said to exhibit interdependent action. However, types 2, 3, 5 and 9 can in fact be observed even when D can be represented as D = $A_0 \bigvee A_1 E_1 \bigvee A_2 \overline{E_1} \bigvee A_3 E_2 \bigvee A_4 \overline{E_2}$. For example, if $A_5 = A_6 = A_7 = A_8 = 0$ but if for some some individual ω , $A_0(\omega) = A_2(\omega) = A_4(\omega) = 0$ and $A_1(\omega) = A_3(\omega) = 1$ so that $D(\omega) = E_1 \bigvee E_2$ then, as seen above, this would give rise to response type 2. Similarly if $A_0(\omega) = A_1(\omega) = A_4(\omega) = 0$ and $A_2(\omega) = A_3(\omega) = 1$ then this would give rise to response type 3; if $A_0(\omega) = A_2(\omega) = A_3(\omega) = 0$ and $A_1(\omega) = A_4(\omega) = 1$ this would give rise to response type 5; if $A_0(\omega) = A_1(\omega) = A_3(\omega) = 0$ and $A_2(\omega) = A_4(\omega) = 1$ this would give rise to response type 9. Response types 2, 3, 5 and 9 might of course also arise from biologic interactions. As noted above, response type 2 would arise if $A_0(\omega) = A_1(\omega) = A_2(\omega) = A_3(\omega) = A_4(\omega) = A_5(\omega) = 0$ and $A_5(\omega) = A_6(\omega) = A_7(\omega) = 1$. Without further information concerning which biological causes are present we cannot, in the case of types 2, 3, 5 and 9, ascertain from the counterfactual response patterns alone whether or not a biologic interaction is present. The types that Greenland and Poole classify as not representing causal interdependence (types 1, 4, 6, 11, 13, 16) can, like types 2, 3, 5 and 9, also all be observed when D can be represented as $D = A_0 \bigvee A_1 E_1 \bigvee A_2 E_1 \bigvee A_3 E_2 \bigvee A_4 E_2$. But all of these types, other than type 16, can also be observed when one or more of A_5, A_6, A_7, A_8 are non-zero. In contrast types 7, 8, 10, 12, 14, and 15 cannot be observed when $A_5 = A_6 = A_7 = A_8 = 0$, i.e. when $D = A_0 \bigvee A_1 E_1 \bigvee A_2 \overline{E_1} \bigvee A_3 E_2 \bigvee A_4 \overline{E_2}$. These six types thus clearly do constitute instances

Research Archive

of biologic interaction because one or more of A_5 , A_6 , A_7 , A_8 must be non-zero for such types to be present. Thus although biologic interactions will sometimes be unidentified even when the counterfactual response patterns for all individuals are known, they will not always be unidentified.

As suggested above and as has been demonstrated in related work (VanderWeele and Robins 2006a) when an individual of type 7, 8, 10, 12, 14, or 15 is present in the population there must be synergism - there is no sufficient cause representation for D in which E_1 and E_2 never appear in the same sufficient cause. We will use the term "definite interdependence," which we make precise in Definition 1, to refer to a counterfactual response pattern which necessarily entails the presence of a biologic interaction.

Note that $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$ if and only if individual ω is response type 7 or 8; also $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{01}(\omega) = 1$ if and only if individual ω is response type 10 or 12; also $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{10}(\omega) = 1$ if and only if individual ω is response type 10 or 14; and finally $D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ if and only if individual ω is response type 7 or 15. The presence of one of the six types that necessarily entails a biologic interaction is thus equivalent to the presence of an individual ω for whom one of the following four conditions hold: $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$; or $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{01}(\omega) = 1$; or $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{00}(\omega) = 1$. Consequently, we define definite interdependence as follows.

DEFINITION 1 (Definite Interdependence). Suppose that D and two of its causes, E_1 and E_2 , are binary. We say that there is definite interdependence between the effect of E_1 and E_2 on D if there exists an individual ω for whom one of the following holds: $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$; or $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{01}(\omega) = 1$; or $D_{11}(\omega) = 0$ and $D_{10}(\omega) = 1$; or $D_{11}(\omega) = 0$ and $D_{00}(\omega) = 1$.

The definition of definite interdependence is equivalent to the presence within a population of an individual with a counterfactual response pattern of type 7, 8, 10, 12, 14, or 15. As defined above, if E_1 and E_2 exhibit definite interdependence then there must be a biologic interaction between E_1 and E_2 . If $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$ then $A_5 \neq 0$ and there will be synergism between E_1 and E_2 . If $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{01}(\omega) = 1$ then $A_6 \neq 0$ and there will be synergism between $\overline{E_1}$ and E_2 . If $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{10}(\omega) = 1$ then $A_7 \neq 0$ and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. If $D_{11}(\omega) = D_{00}(\omega) = 0$ and $D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ then $A_8 \neq 0$ and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. If $D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ then $A_8 \neq 0$ and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. If $D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ then $A_8 \neq 0$ and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. If $D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = 1$ then $A_8 \neq 0$ and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. As made clear in the discussion above, however, although definite interdependence is sufficient for a biologic interaction, it is not necessary. There may be a biologic interaction between E_1 and E_2 even if they do not exhibit definite interdependence.

Greenland and Poole (1988) note that there is a one-to-one correspondence between response types 8, 12, 14 and 15 and "cause types" corresponding to $A_5(\omega) = 1$, $A_6(\omega) = 1$, $A_7(\omega) = 1$ and $A_8(\omega) = 1$ respectively with all other $A_i(\omega) = 0$. They also note that response type 16 arises if and only if $A_i(\omega) = 0$ for all *i*. However, they claim that there are no other one-to-one correspondences for the remaining 11 response types. They fail to notice that response type 7 arises if and only if $A_5(\omega) = 1$ and $A_8(\omega) = 1$ with $A_i(\omega) = 0$ for all $i \notin \{5, 8\}$ and that response type 10 arises if and only if $A_6(\omega) = 1$ and $A_7(\omega) = 1$ with $A_i(\omega) = 0$ for all $i \notin \{6, 7\}$. We will see below that this insight that response types 7 and 10 necessarily entail a biologic interaction is important in constructing statistical tests for the presence of a biologic interaction. A further comment relating definite interdependence and what Greenland and Poole (1988) define as causal interdependence is given in the Appendix.

The definition of definite interdependence given above is invariant to the relabeling of the levels of E_1 and E_2 i.e. relabeling for E_1 and/or for E_2 the level "1" as "0" and "0" as "1." Definite interdependence as defined above is not however invariant to the relabeling of the levels of D. If D is relabeled so that "1" is "0" and "0" is "1" then types 8, 12, 14, and 15 become types 9, 5, 3, and 2 respectively and these latter types do not exhibit definite interdependence. We argue, however, that the fact that definite interdependence is not invariant to the relabeling of the levels

of D is actually in accord with intuition. As above, consider again the presence in an individual of a certain dominant genetic trait D. Let E_1 denote the gene inherited from the mother and E_2 the gene inherited from the father. We would in this case represent the sufficient causes for D by $D = E_1 \bigvee E_2$. No biologic interaction or interdependence between E_1 or E_2 would be thought to be present. If, on the other hand, the outcome of interest were the recessive trait \overline{D} then \overline{D} would be present if and only if both the gene inherited from the mother and the gene inherited from the father were recessive, $\overline{E_1}$ and $\overline{E_2}$, respectively. We would then represent the sufficient causes for \overline{D} by $\overline{D} = \overline{E_1}\overline{E_2}$ and thus for \overline{D} , $\overline{E_1}$ and $\overline{E_2}$ would be thought to have a biologic interaction. The presence of a biologic interaction for an outcome does not imply the presence of a biologic interaction for the complement of that outcome.

3. Testing for General Biologic Interactions

When there is no confounding of the causal effects of E_1 and E_2 on D or if there exists a set of variables C such that conditioning on C suffices to control for the confounding of the causal effects of E_1 and E_2 on D then it is possible to develop statistical tests for the presence of a biologic interaction. Theorem 1 gives a condition which is sufficient for the presence of a biologic interaction and which with data can be statistically tested. The proof of Theorem 1 and that of Theorem 2 below are given in the Appendix.

THEOREM 1. Suppose that D and two of its causes, E_1 and E_2 , are binary. Let C be a set of variables that suffices to control for the confounding of the causal effects of E_1 and E_2 on D then if for any value c of C we have that $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) > 0$ then E_1 and E_2 have a biologic interaction.

When the condition of Theorem 1 is met, an individual of either type 7 or type 8 must be present and from the discussion above it follows that there must be synergism between E_1 and E_2 . Theorem 1 has analogues for testing for synergism between E_1 and $\overline{E_2}$ or between $\overline{E_1}$ and E_2 or between $\overline{E_1}$ and $\overline{E_2}$. If for some c, $P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) > 0$ then an individual of type 10 or type 14 must be present and there will be synergism between E_1 and $\overline{E_2}$. If $P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) > 0$ then an individual of type 10 or type 10 or type 12 must be present and there will be synergism between $\overline{E_1}$ and E_2 . If $P(D = 1|E_1 = 0, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) > 0$ then an individual of type 7 or type 15 must be present and there will be synergism between $\overline{E_1}$ and $\overline{E_2}$. Theorem 1 and its analogues demonstrate that the claim of Rothman and Greenland (1998, p. 339) that inference about the presence of biologic interactions "must make untestable assumptions about the absence of other response types" is false. Rothman and Greenland failed to observe the implication given in Theorem 1. Theorem 1 makes no assumption about the absence of any response type.

In general to test the null that $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) \le 0$ for a particular sample one may let n_{ij} denote the number of individuals in stratum C = c with $E_1 = i$ and $E_2 = j$ and let let d_{ij} denote the number of individuals in stratum C = c with $E_1 = i$, $E_2 = j$ and D = 1 then tests of the null hypothesis $P(D = 1|E_1 = 1, E_2 = 0, C = c) \le 0$ can be constructed using critical regions of the following form: $\{\frac{\binom{d_{11}}{n_{11}} - \frac{d_{10}}{n_{01}} - \frac{d_{10}}{n_{10}}}{\binom{d_{11}(n_{11} - d_{11})}{n_{01}^2} + \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3}} > \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3} = \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3}}$

 $Z_{1-\alpha}$ }. This can be seen by letting p_{ij} denote the true probability of D = 1 conditional on $E_1 = i, E_2 = j$ and C = c. The hypothesis $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) \le 0$ is that $(p_{11} - p_{01} - p_{10}) \le 0$. We have that $d_{ij} \sim Bin(n_{ij}, p_{ij})$ with $\mathbb{E}[\frac{d_{ij}}{n_{ij}}] = p_{ij}$ and $Var(\frac{d_{ij}}{n_{ij}}) = \frac{p_{ij}(1-p_{ij})}{n_{ij}}$. By the central limit

central limit theorem $\frac{(\frac{d_{11}}{n_{11}} - \frac{d_{01}}{n_{10}} - \frac{d_{10}}{n_{10}}) - (p_{11} - p_{01} - p_{10})}{\sqrt{\frac{p_{11}(1 - p_{11})}{n_{11}} + \frac{p_{01}(1 - p_{01})}{n_{01}} + \frac{p_{10}(1 - p_{10})}{n_{10}}}} \sim N(0, 1) \text{ and by Slutsky's theorem we have}$

$$\frac{\left(\frac{d_{11}}{n_{11}} - \frac{d_{01}}{n_{01}} - \frac{d_{10}}{n_{10}}\right) - (p_{11} - p_{01} - p_{10})}{\sqrt{\frac{d_{11}(n_{11} - d_{11})}{n_{11}^3} + \frac{d_{01}(n_{01} - d_{01})}{n_{01}^3} + \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3}} \sim N(0, 1).$$
 To test the hypothesis $H_0: (p_{11} - p_{01} - p_{10}) \le 0$ vs.

 $H_A: (p_{11} - p_{01} - p_{10}) > 0 \text{ one may thus use the test statistic: } \frac{\left(\frac{d_{11}}{n_{11}} - \frac{d_{01}}{n_{01}} - \frac{d_{10}}{n_{10}}\right)}{\sqrt{\frac{d_{11}(n_{11} - d_{11})}{n_{11}^3} + \frac{d_{01}(n_{01} - d_{01})}{n_{01}^3} + \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3}}}}$

If C consists of a small number of binary or categorical variables then it may be possible to use the tests constructed above to test all strata of C. When C includes a continuous variable or many binary and categorical variables such testing becomes difficult because the data in certain strata of C will be sparse. One might then model the conditional probabilities $P(D = 1|E_1, E_2, C)$ using a binomial or Poisson regression model with a linear link (Greenland, 1991; Wacholder, 1986; Zou, 2004; Greenland, 2004; Spiegelman and Hertzmark, 2005). For case-control studies it will be necessary to use an adapted set of modeling techniques (Wild, 1991; Wacholder, 1996; Greenland, 2004).

4. Testing for Biologic Interactions under the Assumption of Monotonic Effects

We next consider a context in which the direction of the effect (positive or negative) that E_1 and E_2 have on D is known. We make these ideas precise by introducing the concept of a monotonic effect. Considerable intuition regarding biologic interactions can be garnered by the consideration of the setting of monotonic effects. Furthermore, as will be seen shortly, the setting of monotonic effects also allows for the construction of more powerful tests for detecting biologic interactions than is possible without the assumption.

DEFINITION 2 (Monotonic Effect): We will say that E_1 has a positive monotonic effect on Dif for all individuals ω we have $D_{ij}(\omega) \geq D_{i'j}(\omega)$ whenever $i \geq i'$; we will say that E_2 has a positive monotonic effect on D if for all individuals ω we have $D_{ij}(\omega) \geq D_{ij'}(\omega)$ whenever $j \geq j'$. Similarly, we will say that E_1 has a negative monotonic effect on D if for all individuals ω we have $D_{ij}(\omega) \leq D_{i'j}(\omega)$ whenever $i \geq i'$ and that E_2 has a negative monotonic effect on D if for all individuals ω we have $D_{ij}(\omega) \leq D_{ij'}(\omega)$ whenever $j \geq j'$.

The definition of a monotonic effect essentially requires that some intervention either increase or decrease some other variable D not merely on average over the entire population but rather for every individual in that population regardless of the other intervention. 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. VanderWeele and Robins (2006b) provide further discussion of the idea of a monotonic effect and relate the concept to causal effects, covariance, confounding and bias.

Theorem 2 gives a result similar to that of Theorem 1 but under the assumption that both E_1 and E_2 have positive monotonic effects on D.

THEOREM 2. Suppose that D and two of its causes, E_1 and E_2 , are binary and that E_1 and E_2 have a positive monotonic effect on D. Let C be a set of variables that suffices to control for the confounding of the causal effects of E_1 and E_2 on D then if for any value c of C we have that $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) > P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c)$ then E_1 and E_2 have a biologic interaction.

The condition provided in Theorem 2 has obvious analogues if one or both of E_1 and E_2 have a negative monotonic effect rather than a positive monotonic effect on D. If the condition of

Theorem 2 is met, an individual of type 8 must be present. Individuals of type 7, the other type that entails synergism between E_1 and E_2 , are precluded because E_1 has a positive monotonic effect on D (and similarly because E_2 has a positive monotonic effect on D). Rothman and Greenland (1998) note the equivalent result in the setting of no confounding factors. The condition of Theorem 2 can be tested in a manner analogous to the condition of Theorem 1 in the previous section. In general to test the null that $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) \leq P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c)$ for a particular sample, tests of the null hypothesis can be constructed using critical regions of the following form: $\frac{d_{11} - d_{01}}{d_{01} - d_{00}}$

 $\{\frac{(\frac{d_{11}}{n_{11}} - \frac{d_{01}}{n_{01}}) - (\frac{d_{10}}{n_{10}} - \frac{d_{00}}{n_{00}})}{\sqrt{\frac{d_{11}(n_{11} - d_{11})}{n_{11}^3} + \frac{d_{01}(n_{01} - d_{01})}{n_{01}^3} + \frac{d_{10}(n_{10} - d_{10})}{n_{10}^3} + \frac{d_{00}(n_{00} - d_{00})}{n_{00}^3}} > Z_{1-\alpha}\}}.$

The general condition of Theorem 1 for detecting the presence of a biologic interaction between E_1 and E_2 , $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) > 0$, is clearly stronger than the condition, $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) > 0$, required in the setting of monotonic effects. Indeed the former clearly implies the latter. The statistical tests based on this condition in the setting of monotonic effects will thus be more powerful than the equivalent tests in the general setting.

5. Examples and Discussion

Theorem 2 suggests the risk difference scale as the means by which to test for a biologic interaction in the presence of monotonic effects. Theorem 2 can be interpreted as stating that if the risk difference for E_1 in the strata $E_2 = 1$ is greater than the risk difference for E_1 in the strata $E_2 = 0$ then E_1 and E_2 have a biologic interaction. The condition can also be re-written as $P(D = 1|E_1 = 1, E_2 =$ $1, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) > \{P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 =$ $0, E_2 = 0, C = c)\} + \{P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c)\}$ i.e. the effect of E_1 and E_2 is greater than the sum of the effects of E_1 and E_2 separately. The result is in many ways intuitive and not at all surprising. Several distinctions between the categories of effect modification on the risk difference scale and that of definite interdependence or biologic interaction must be kept in mind however as the following examples make clear. First, Example 1 demonstrates that it is possible to have effect modification on the risk difference scale with no biologic interaction. This may arise when the effect modification is in the opposite direction of that required by Theorem 2.

EXAMPLE 1. Suppose that D, E_1 and E_2 are binary and that $D = A_0 \bigvee A_1 E_1 \bigvee A_2 E_2$ so that E_1 and E_2 have a positive monotonic effect on D and E_1 and E_2 do not have a biologic interaction. Suppose further that the causal effects of E_1 and E_2 on D are unconfounded. Let $P(A_0) = a_0, P(A_1) =$ $a_1, P(A_2) = a_2, P(A_0A_1) = a_{01}, P(A_0A_2) = a_{02}, P(A_1A_2) = a_{12}, P(A_0A_1A_2) = a_{012}.$ We then have $P(D = 1|E_1 = 0, E_2 = 0) = P(A_0) = a_0; P(D = 1|E_1 = 1, E_2 = 0) = P(A_0 \bigvee A_1) = a_0 + a_1 - a_{01};$ $P(D = 1|E_1 = 0, E_2 = 1) = P(A_0 \bigvee A_2) = a_0 + a_2 - a_{02}; \text{ and } P(D = 1|E_1 = 1, E_2 = 1) = 0$ $P(A_0 \bigvee A_1 \bigvee A_2) = a_0 + a_1 + a_2 - a_{01} - a_{02} - a_{12} + a_{012}$. Conditional on $E_2 = 0$, the risk difference for E_1 is given by: $P(D = 1 | E_1 = 1, E_2 = 0) - P(D = 1 | E_1 = 0, E_2 = 0) = a_0 + a_1 - a_{01} - a_0 = a_1 - a_{01}$. Conditional on $E_2 = 1$, the risk difference for E_1 is given by: $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1)$ $1|E_1 = 0, E_2 = 1) = a_0 + a_1 + a_2 - a_{01} - a_{02} - a_{12} + a_{012} - (a_0 + a_2 - a_{02}) = (a_1 - a_{01}) - (a_{12} - a_{012}).$ In this example, $P(D = 1 | E_1 = 1, E_2 = 1) - P(D = 1 | E_1 = 0, E_2 = 1) = (a_1 - a_{01}) - (a_{12} - a_{012}) \neq 0$ $a_1 - a_{01} = P(D = 1 | E_1 = 1, E_2 = 0) - P(D = 1 | E_1 = 0, E_2 = 0)$. We see then from this example that we can have effect modification on the risk difference scale ("statistical interaction") even when no biologic interaction is present. This will occur whenever $(a_{12} - a_{012}) \neq 0$ i.e. when $P(A_1A_2) \neq P(A_0A_1A_2)$ or equivalently $P(A_0 = 1 | A_1 = 1, A_2 = 1) < 1$. This example is consider further in the Appendix to relate biologic interactions to the multiplicative survival model.

Example 2 below demonstrates that the absence of effect modification on the risk difference scale does not imply the absence of a biologic interaction. In other words, if $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) > P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c)$ then there must be a biologic interaction but even if $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) \le P(D = 1|E_1 = 1, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) + P(D = 1|E_1 = 0, E_2 = 0, C = c) +$

EXAMPLE 2. Suppose that D, E_1 and E_2 are binary, that E_1 and E_2 are independent and that $D = A_0 \bigvee A_1 E_1 \bigvee A_2 E_2 \bigvee A_3 E_1 E_2$ so that E_1 and E_2 have a positive monotonic effect on D and E_1 and E_2 do have a biologic interaction. Suppose further that the causal effects of E_1 and E_2 on D are unconfounded. Let $P(A_0) = a_0, P(A_1) = a_1, P(A_2) = a_2, P(A_3) = a_3, P(A_0A_1) = a_{01}, P(A_0A_2) = a_{01}, P(A_0A_2)$ $a_{02}, ..., P(A_0A_1A_2A_3) = a_{0123}$. We then have $P(D = 1|E_1 = 0, E_2 = 0) = P(A_0) = a_0$; $P(D = a_0) = P(A_0) = a_0$; $P(A_0) =$ $1|E_1 = 1, E_2 = 0) = P(A_0 \bigvee A_1) = a_0 + a_1 - a_{01}; P(D = 1|E_1 = 0, E_2 = 1) = P(A_0 \bigvee A_2) = 0$ $a_0 + a_2 - a_{02}$; and $P(D = 1 | E_1 = 1, E_2 = 1) = P(A_0 \bigvee A_1 \bigvee A_2) = (a_0 + a_1 + a_2 + a_3) - (a_{01} + a_{02} + a_{03} + a_{12} + a_{13} + a_{23}) + (a_{012} + a_{013} + a_{023} + a_{123}) - a_{0123}$. Thus $P(D = 1 | E_1 = 1, E_2 = 1) = P(A_0 \bigvee A_1 \bigvee A_2) = (a_0 + a_1 + a_2 + a_3) - (a_{01} + a_{02} + a_{03} + a_{12} + a_{13} + a_{23}) + (a_{012} + a_{013} + a_{023} + a_{123}) - a_{0123}$. $1) - P(D = 1|E_1 = 0, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 0) + P(D = 1|E_1 = 0, E_2 = 0) =$ $(a_{012} - a_{12}) + a_3 - (a_{03} + a_{13} + a_{23}) + (a_{013} + a_{023} + a_{123}) - a_{0123}$. Suppose now that with probability 0.5, $A_0 = 0, A_1 = 0, A_2 = 0, A_3 = 1$ and with probability 0.5, $A_0 = 0, A_1 = 1, A_2 = 1, A_3 = 0$ so that $a_3 = 0.5$ and $a_{12} = 0.5$ and $a_{012} = a_{03} = a_{13} = a_{23} = a_{013} = a_{023} = a_{123} = a_{0123} = 0$ then $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 0) + P(D = 1|E_1 = 1, E_2 = 0) + P(D = 1|E_1 = 1, E_2 = 0) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1, E_2 = 1) + P(D = 1|E_1 = 1) + P(D =$ $1|E_1 = 0, E_2 = 0| = a_3 - a_{12} = 0.5 - 0.5 = 0$ and so although a biologic interaction is present the inequality $P(D = 1 | E_1 = 1, E_2 = 1) - P(D = 1 | E_1 = 0, E_2 = 1) > P(D = 1 | E_1 = 1, E_2 = 1)$ $(0) - P(D = 1|E_1 = 0, E_2 = 0)$ fails to hold. The example demonstrates that although the inequality is a sufficient condition for a biologic interaction under the setting of monotonic effects, it is not necessary. It is also interesting to note that in this example $P(D = 1|E_1 = 1, E_2 =$ $1) - P(D = 1|E_1 = 0, E_2 = 1) - P(D = 1|E_1 = 1, E_2 = 0) + P(D = 1|E_1 = 0, E_2 = 0) =$ $\{a_3 - (a_{03} + a_{13} + a_{23}) + (a_{013} + a_{023} + a_{123}) - a_{0123}\} - (a_{12} - a_{012})$ and this final expression can be rewritten as $P(A_3\overline{A}_0\overline{A}_1\overline{A}_2) - P(A_1A_2\overline{A}_0)$ suggesting that the more likely that A_3 occurs when A_0, A_1, A_2 are absent, the more power the test implied by Theorem 2 will have to detect the biologic interaction; on the other hand the more likely that A_1 and A_2 occur together when A_0 is absent, the less power the test implied by Theorem 2 will have to detect the biologic interaction.

The contrast between Examples 1 and 2 is interesting. Example 1 demonstrated that effect modification could be present without a biologic interaction. In Example 1, effect modification on the risk difference scale would be present whenever $P(A_1A_2) \neq P(A_0A_1A_2)$ suggesting that, in general, effective modification on the risk difference scale may be present without a biologic interaction if the various background causes A_0 , A_1 and A_2 can occur simultaneously i.e. when multiple causal mechanisms may be simultaneously operative. It is, of course, also possible to have effect modification that is attributable solely to biologic interactions rather than to the background causes. Example 2 considered the general case of a biologic interaction between E_1 and E_2 under the setting of monotonic effects. The expression for $\{P(D=1|E_1=1,E_2=1)-P(D=1|E_1=1)\}$ $(0, E_2 = 1)$ - { $P(D = 1 | E_1 = 1, E_2 = 0) - P(D = 1 | E_1 = 0, E_2 = 0)$ } could be rewritten as $(a_{012} - a_{12}) + (a_3 - a_{03} - a_{13} - a_{23} + a_{013} + a_{023} + a_{123} - a_{0123})$. For no effect modification on the risk difference scale to be present in Example 2 the sum of these two terms would have to be zero. Note that each part of the second term involves the subscript 3. The second term can thus be seen as the synergistic component; it will be zero when $A_3 = 0$. We saw in Example 1 that the first term being zero, $(a_{012} - a_{12}) = 0$, was the condition for no effect modification in the case of $A_3 = 0$. Suppose that $(a_{012} - a_{12}) = 0$ but $A_3 \neq 0$ and $(a_3 - a_{03} - a_{13} - a_{23} + a_{013} + a_{023} + a_{123} - a_{0123}) \neq 0$ then the effect modification in Example 2 would be attributable solely to the biologic interaction (i.e. no effect modification would be present if $A_3 = 0$). Thus in Example 1, the effect modification was wholly

attributable to the possibility of the background causes A_0 , A_1 and A_2 occurring simultaneously and in Example 2, if $(a_{012} - a_{12}) = 0$, the effect modification would be wholly attributable to the presence of a biologic interaction. In general, effect modification may arise either due to background causes or due to the presence of biologic interactions or due to both.

Example 3 demonstrates that if it is not the case that both E_1 and E_2 have a monotonic effect on D then we may have $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) > P(D = 1|E_1 = 1, E_2 = 0) - P(D = 1|E_1 = 0, E_2 = 0)$ even when there is no biologic interaction and that in such cases we can also have E_2 acting as a qualitative effect modifier for the risk difference of E_1 on Dwithout E_1 and E_2 having a biologic interaction.

EXAMPLE 3. Suppose that D, E_1 and E_2 are binary, that E_1 and E_2 are independent and that $D = A_1E_1 \bigvee A_2\overline{E_1} \bigvee A_3\overline{E_2}$ so that E_1 and E_2 do not have a biologic interaction. Suppose further that the causal effects of E_1 and E_2 on D are unconfounded. Finally, suppose that with probability 0.3, $A_1 = 1, A_2 = 0, A_3 = 1$; with probability 0.3, $A_1 = 1, A_2 = 0, A_3 = 0$; and with probability 0.4, $A_1 = 0, A_2 = 1, A_3 = 0$ so that $a_1 = 0.6, a_2 = 0.4, a_3 = 0.3, a_{13} = 0.3$ and $a_{23} = 0$. We then have $P(D = 1|E_1 = 0, E_2 = 0) = P(A_2 \bigvee A_3) = a_2 + a_3 - a_{23}$; $P(D = 1|E_1 = 1, E_2 = 0) = P(A_1 \bigvee A_3) = a_1 + a_3 - a_{13}$; $P(D = 1|E_1 = 0, E_2 = 1) = P(A_2) = a_2$; and $P(D = 1|E_1 = 1, E_2 = 1) = P(A_1) = a_1$. Conditional on $E_2 = 0$, the risk difference for E_1 is given by: $P(D = 1|E_1 = 1, E_2 = 0) - P(D = 1|E_1 = 0, E_2 = 0) = a_1 + a_3 - a_{13} - (a_2 + a_3 - a_{23}) = a_1 - a_2 - a_{13} + a_{23} = 0.6 - 0.4 - 0.3 = -0.1$. Conditional on $E_2 = 1$, the risk difference for E_1 is given by: $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) = a_1 - a_2 = 0.6 - 0.4 = 0.2$. In this example, $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) > P(D = 1|E_1 = 1, E_2 = 0) - P(D = 1|E_1 = 0, E_2 = 1) = a_1 - a_2 = 0.6 - 0.4 = 0.2$. In this example, $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) > P(D = 1|E_1 = 1, E_2 = 0) - P(D = 1|E_1 = 0, E_2 = 1) = a_1 - a_2 = 0.6 - 0.4 = 0.2$. In this example, $P(D = 1|E_1 = 1, E_2 = 1) - P(D = 1|E_1 = 0, E_2 = 1) > P(D = 1|E_1 = 1, E_2 = 0) - P(D = 1|E_1 = 0, E_2 = 0)$ but no biologic interaction was present. We see also from this example that we can have qualitative effect modification even when no biologic interaction is present.

The three examples above clarify the conceptual distinction between effect modification on the risk difference scale and biologic interactions, even in the presence of monotonic effects. There can be effect modification on the risk difference scale without the presence of a biologic interaction. There can be a biologic interaction without the risk difference condition $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) > P(D = 1|E_1 = 1, E_2 = 0, C = c) - P(D = 1|E_1 = 0, E_2 = 0, C = c)$ holding. And, finally, outside the context of monotonic effects, we may have $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) > P(D = 1|E_1 = 1, E_2 = 0, C = c)$ without the presence of a biologic interaction.



Appendix.

Proof of Theorem 1.

Suppose that for some set of variables V, $\mathbb{E}[D_{11} - D_{01} - D_{10}|V = v] > 0$ then there must be some individual ω for whom V = v and $D_{11}(\omega) = 1$ but $D_{01}(\omega) = D_{10}(\omega) = 0$ for if one of $D_{01}(\omega), D_{10}(\omega)$ were always 1 whenever $D_{11}(\omega) = 1$ then $D_{11}(\omega) - D_{01}(\omega) - D_{10}(\omega)$ would be less than or equal to zero for all ω and so we would have that $\mathbb{E}[D_{11} - D_{01} - D_{10}|V = v] \leq 0$. Let V = C, definite interdependence and thus a biologic interaction is implied by the condition $\mathbb{E}[D_{11} - D_{01} - D_{10}|C = c] > 0$. Because C is a set of variables that suffices to control for the confounding of the causal effects of E_1 and E_2 on D we have that the counterfactual variables D_{ij} are conditionally independent of (E_1, E_2) given C and so we have, $\mathbb{E}[D_{11} - D_{01} - D_{10}|C = c] = \mathbb{E}[D_{11}|E_1 = 1, E_2 = 1, C = c] - \mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, C = c] = P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c)$. Thus if $P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 0, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 0, C = c) > 0$ then E_1 and E_2 have a biologic interaction.

Proof of Theorem 2.

We first show that if for some set of variables V, $\mathbb{E}[(D_{11} - D_{01}) - (D_{10} - D_{00})|V = v] > 0$ for some v then there must be a biologic interaction. For each individual ω define $B_0(\omega)$, $B_1(\omega)$, $B_2(\omega)$ and $B_3(\omega)$ as follows: $B_0(\omega) = 1$ if $D_{00}(\omega) = 1$ and 0 otherwise; $B_1(\omega) = 1$ if $D_{10}(\omega) = 1$ and 0 otherwise; $B_2(\omega) = 1$ if $D_{01}(\omega) = 1$ and 0 otherwise; and $B_3(\omega) = 1$ if $D_{11}(\omega) = 1$ and $D_{10}(\omega) = D_{01}(\omega) = 0$ and 0 otherwise. Then $D_{00} = B_0$, $D_{10} = B_0 \bigvee B_1$, $D_{01} = B_0 \bigvee B_2$, $D_{11} = B_0 \bigvee B_2$, $D_{11} = B_0 \bigvee B_0$ $B_0 \bigvee B_1 \bigvee B_2 \bigvee B_3$. Suppose there is no biologic interaction; then $B_3(\omega) = 0$ for all $\omega \in \Omega$ so that $D_{11} = B_0 \bigvee B_1 \bigvee B_2$. Let $P(B_0|V = v) = b_0^v$, $P(B_1|V = v) = b_1^v$, $P(B_2|V = v) = b_2^v$, $P(B_0B_1|V=v) = b_{01}^v, P(B_0B_2|V=v) = b_{02}^v, P(B_1B_2|V=v) = b_{12}^v$ and $P(B_0B_1B_2|V=v) = b_{012}^v$. Then $P(B_0|V=v) = b_0^v$; $P(B_0 \bigvee B_1|V=v) = b_0^v + b_1^v - b_{01}^v$; $P(B_0 \bigvee B_2|V=v) = b_0^v + b_2^v - b_{02}^v$;
$$\begin{split} P(B_0 \bigvee B_1 \bigvee B_2 | V = v) &= b_0^v + b_1^v + b_2^v - (b_{01}^v + b_{02}^v + b_{12}^v) + b_{012}^v. \quad \mathbb{E}[(D_{11} - D_{01}) - (D_{10} - D_{00})|V = v] \\ v] &= \{P(B_0 \bigvee B_1 \bigvee B_2 | V = v) - P(B_0 \bigvee B_2 | V = v)\} - \{P(B_0 \bigvee B_1 | V = v) - P(B_0 | V = v)\} = \{P(B_0 \bigvee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \bigvee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} - \{P(B_0 \vee B_1 | V = v) - P(B_0 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 \vee B_2 | V = v)\} \\ = \{P(B_0 \vee B_1 \vee B_2 | V = v) - P(B_0 \vee B_2 \vee B_$$
 $[\{b_0^v + b_1^v + b_2^v - (b_{01}^v + b_{02}^v + b_{12}^v) + b_{012}^v\} - \{b_0^v + b_2^v - b_{02}^v\}] - [\{b_0^v + b_1^v - b_{01}^v\} - b_0^v] = (b_{012}^v - b_{12}^v) + b_{12}^v - b_{01}^v) - (b_{11}^v - b_{01}^v) = b_{012}^v - b_{12}^v < 0.$ Thus if $\mathbb{E}[(D_{11} - D_{01}) - (D_{10} - D_{00})|V = v] > 0$ we cannot have $B_3(\omega) = 0$ for all ω and so there must be a biologic interaction of the effect of E_1 and E_2 on D. Now let V = C then we have that a biologic interaction is implied by the condition $\mathbb{E}[(D_{11} - D_{01}) - (D_{10} - D_{00})|C = c] > 0$. Because C is a set of variables that suffices to control for the confounding of the causal effects of E_1 and E_2 on D we have that the counterfactual variables D_{ij} are conditionally independent of (E_1, E_2) given C and so we have, $\mathbb{E}[(D_{11} - D_{01}) - (D_{10} - D_{00})|C = c] = \{\mathbb{E}[D_{11}|C = c] - \mathbb{E}[D_{01}|C = c]\} - \{\mathbb{E}[D_{10}|C = c] - \mathbb{E}[D_{00}|C = c]\} - \mathbb{E}[D_{00}|C = c]\} - \mathbb{E}[D_{00}|C = c] - \mathbb{E}[D_{00}|C = c] - \mathbb{E}[D_{00}|C = c]\} - \mathbb{E}[D_{00}|C = c] - \mathbb{E}[D_{00}|$ $c])\} = \{\mathbb{E}[D_{11}|E_1 = 1, E_2 = 1, C = c] - \mathbb{E}[D_{01}|E_1 = 0, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, E_1 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 0, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 1, C = c]\} - \{\mathbb{E}[D_{10}|E_1 = 1, E_2 = 1, C = c]\}$ $0, C = c] - \mathbb{E}[D_{00}|E_1 = 0, E_2 = 0, C = c]) = \{P(D = 1|E_1 = 1, E_2 = 1, C = c) - P(D = 1|E_1 = 1, E_2 = 1, C = c)\}$ $0, E_2 = 1, C = c$ } - { $P(D = 1 | E_1 = 1, E_2 = 0, C = c$ } - $P(D = 1 | E_1 = 0, E_2 = 0, C = c$ }. Thus if $P(D=1|E_1=1,E_2=1,C=c) - P(D=1|E_1=0,E_2=1,C=c) > P(D=1|E_1=1,E_2=0,C=c) = P(D=1|E_1=1,E_1=0,C=c) = P(D=1|E_1=1,E_1=0,C=c) = P(D=1|E_1=1,E_1=0,C=c) = P(D=1|E_1=1,E_1=0,C$ $c) - P(D = 1 | E_1 = 0, E_2 = 0, C = c)$ then E_1 and E_2 must have a biologic interaction.

Definite Interdependence and Causal Interdependence

Greenland and Poole (1988) use "causal interdependence" to refer to the presence of an individual of response type 2, 3, 5, 7, 8, 9, 10, 12, 14, or 15. We have not included response types 2, 3, 5 and 9 in our definition of "definite interdependence." As noted above, response types 2, 3, 5 and 9 may arise without the presence of a biologic interaction. Now suppose that E_1 and E_2 are gun shots from two snipers who fire at a particular individual at times 1 and 2 respectively. If both snipers are perfectly accurate and D is the outcome death then the counterfactual response pattern for the individual at whom the shots are being fired is type 2. In this case there is no "biologic interaction" between the gun shots of the two snipers; one might however say that the the two gun shots are "causally interdependent" because if the individual is killed by bullet E_1 he cannot be killed by bullet E_2 and

vice versa. However, although counterfactual response patterns of types 2, 3, 5 and 9 can arise from two causes which are in some sense interdependent, these types can also arise in a context in which the two causes cannot be conceived of as being interdependent. Consider for example the presence in an individual of a certain dominant genetic trait denoted by phenotype D so that D denotes the absence of this dominant trait. Let E_1 denote the gene inherited from the mother and E_2 the gene inherited from the father. If either E_1 or E_2 or both E_1 and E_2 take values corresponding to the dominant gene, the child will have the dominant trait. The presence of the dominant genetic trait D is of counterfactual response type 2; but in this case the causes E_1 and E_2 cannot be conceived of as being in any sense interdependent. If both E_1 and E_2 take values corresponding to the dominant gene, the effect of E_1 on D does not in this example, unlike the sniper example, preclude E_2 from also having an effect on D. We see then that individuals of counterfactual types 2, 3, 5 and 9 can arise with or without the presence of a biologic interaction and with or without interdependence between the two causes E_1 and E_2 . Counterfactual response types 7, 8, 10, 12, 14 and 15 necessarily entail a biologic interaction and thus also a form of interdependence; we have thus defined the class of response types 7, 8, 10, 12, 14 and 15 as those which exhibit "definite interdependence." With regard to other response types, determining whether the two causes are in any way interdependent will require some knowledge of the context and causal mechanisms involved.

Multiplicative Survival Model and Biologic Interactions

Example 1 also sheds light on the conditions under which a multiplicative survival model can be used to test for biologic interactions. The multiplicative survival model is said to hold when $P(D = 0|E_1 = 1, E_2 = 1)P(D = 0|E_1 = 0, E_2 = 0) - P(D = 0|E_1 = 1, E_2 = 0)P(D = 0|E_1 = 0, E_2 = 0)P(D = 0|E_1 = 0,$ $0, E_2 = 1$). In Example 1, the probabilities of survival are: $P(D = 0 | E_1 = 0, E_2 = 0) = 1 - a_0$; $P(D = 0|E_1 = 0, E_2 = 1) = 1 - a_0 - a_1 + a_{01}; P(D = 0|E_1 = 0, E_2 = 1) = 1 - a_0 - a_2 + a_{02};$ $P(D = 0|E_1 = 1, E_2 = 1) = 1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}$ and thus: $P(D = 0|E_1 = 1) = 1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}$ $1, E_2 = 1) \hat{P}(D = 0 | E_1 = 0, E_2 = 0) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012}) = (1 - a_0)(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{01} + a_{01} + a_{02} + a_{01} + a_{01} + a_{02} + a_{01} + a_{01}$ $1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012} - a_0(1 - a_0 - a_1 - a_2 + a_{01} + a_{02} + a_{12} - a_{012});$ but $P(D = 0|E_1 = 1, E_2 = 0)P(D = 0|E_1 = 0, E_2 = 1) = (1 - a_0 - a_1 + a_{01})(1 - a_0 - a_2 + a_{02}) = 0$ $1 - a_0 - a_1 + a_{01} - a_0 - a_2 + a_{02} + a_0^2 + a_0 a_2 - a_0 a_{02} + a_0 a_1 + a_1 a_2 - a_1 a_{02} - a_0 a_{01} - a_2 a_{01} + a_{01} a_{02};$ Thus, $P(D = 0|E_1 = 1, E_2 = 1)P(D = 0|E_1 = 0, E_2 = 0) - P(D = 0|E_1 = 1, E_2 = 0)P(D = 0|E_1 = 0, E_2 = 0)$ $0, E_2 = 1) = (a_{12} - a_1 a_2) - (a_{012} - a_1 a_{02}) - (a_0 a_{12} - a_2 a_{01}) + (a_0 a_{012} - a_{01} a_{02}) \neq 0$ which will generally be non-zero so the multiplicative survival model will fail to hold in this example. However, if A_0 , A_1 and A_2 were independently distributed then the above expression is zero and the multiplicative survival model holds. Somewhat more generally, if A_1 and A_2 were independent of one another and also either A_1 or A_2 were independent of A_0 then the expression would again be zero and the multiplicative survival model would hold. Greenland and Poole (1988) proposed the multiplicative survival model as a means to assess the interdependence versus the independence of causal effects under the setting that the "effects of exposures are probabilistically independent of any background causes, as well as of one another's effect." Example 1 underscores the necessity for the background causes to also be independent of one another when using the multiplicative survival model to detect biologic interactions. More precisely, we have shown that if E_1 and E_2 have a positive monotonic effect on D and if A_1 and A_2 are independent of one another and either A_1 or A_2 is independent of A_0 then the multiplicative survival model will hold when there is no biologic interaction. Therefore, if, under these assumptions, the multiplicative survival model does not hold then one could include that there was a biologic interaction present between E_1 and E_2 .



References

BLOT W.J. AND DAY N.E. (1979). Synergism and interaction: are they equivalent? Am. J. Epidemiol. 100, 99-100.

FLANDERS, D. (2006). Sufficient-component cause and potential outcome models. Working Paper.

GREENLAND, S. (1991). Estimating standardized parameters from generalized linear models. *Stat. Med.* **10**, 1069-1074.

GREENLAND, S. (2004). Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am. J. Epidemiol.* **160**, 301-305.

GREENLAND, S. AND BRUMBACK, B. (2002). An overview of relations among causal modelling methods. *Int. J. Epidemiol.* **31** 1030-1037.

GREENLAND, S. AND POOLE, C. (1988). Invariants and noninvariants in the concept of interdependent effects. *Scand. J. Work Environ. Health* 14, 125-129.

KOOPMAN, J.S. (1981). Interaction between discrete causes. Am. J. Epidemiol. 113, 716-724.

MIETTINEN, O.S. (1982). Causal and preventive interdependence: Elementary principles. *Scand.* J. Work Environ. Health 8, 159-168.

MIETTINEN, O.S. (1985). Modern Epidemiology. New York: John Wiley.

ROTHMAN, K.J. (1976). Causes. Am. J. Epidemiol. 104, 587-592.

ROTHMAN, K.J. AND GREENLAND, S. (1998). Modern Epidemiology. Philadelphia, PA: Lippincott-Raven.

ROTHMAN, K.J., GREENLAND, S. AND WALKER A.M. (1980). Concepts of interaction. Am. J. Epidemiol. 112, 467-470.

SARACCI, R. (1980). Interaction and synergism. Am. J. Epidemiol. 112, 465-466.

SPIEGELMAN, D. AND HERTZMARK, E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. Am. J. Epidemiol. 162, 199-200.

VANDERWEELE, T.J. AND ROBINS, J.M. (2006a). Minimal sufficient causation and directed acyclic graphs. *Ph.D. Thesis*, Chapter 2.

VANDERWEELE, T.J. AND ROBINS, J.M. (2006b). Signed directed acyclic graphs for causal inference. *Ph.D. Thesis*, Chapter 1.

WACHOLDER, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. Am. J. Epidemiol. 123, 174-184.

WACHOLDER, S. (1996). The case-control study as data missing by design: Estimating risk differences. *Epidemiol.* **7**, 145-150.

WILD, C.J. (1991). Fitting prospective regression models to case-control data. *Biometrika* 78, 705-717.

ZOU, G. (2004). A modified Poisson regression approach to prospective studies with binary data. Am. J. Epidemiol. **159**, 702-706.