

Empirical and counterfactual conditions for sufficient cause interactions

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

SUMMARY

Sufficient-component causes are discussed within the deterministic potential outcomes framework so as to formalize notions of sufficient causes, synergism and sufficient cause interactions. Doing so allows for the derivation of counterfactual and empirical conditions for detecting the presence of sufficient cause interactions. The conditions are novel in that, unlike other conditions in the literature, they make no assumptions about monotonicity. The conditions can also be generalized and the conditions for three-way sufficient cause interactions are given explicitly. The statistical tests derived for sufficient cause interactions are compared with and contrasted to interaction terms in standard statistical models.

Some key words: Causal inference; Counterfactual; Interaction; Risk difference; Potential outcome; Sufficient cause; Synergism.

1. INTRODUCTION

Interaction terms in statistical models are frequently used to assess whether or not effects are interdependent. However, whether or not two variables have a statistical interaction may depend on which statistical model is being used (Mantel et al., 1977). When the causes and the outcome under consideration are binary, it has been argued that there is a natural way in which to assess interdependent effects based on a sufficient-component cause framework (Rothman, 1976; Koopman, 1981). This framework makes reference to the actual causal mechanisms involved in bringing about the outcome; when two or more binary causes participate in the same causal mechanism it becomes proper to speak of synergism. In this paper we contribute theory to relate the sufficient-component cause framework to the potential outcomes framework and derive various conditions to statistically test for the presence of sufficient cause interactions. The conditions are novel in that they make no assumption about monotonicity; the conditions can also be generalized to n -way interactions both with and without monotonicity assumptions.

2. SUFFICIENT CAUSES

Two broad conceptualizations of causality can be discerned in the literature, both within philosophy and within statistics and epidemiology. The first may be characterized as giving an account of the effects of particular causes or interventions. In both philosophy and statistics the work is associated with counterfactuals or potential outcomes (Hume, 1748; Neyman, 1923; Lewis, 1973a, 1973b; Rubin, 1974, 1978; Robins, 1986, 1987). The counterfactual or potential outcomes framework has been used extensively in statistics. In contrast, the second conceptualization of causality has received comparatively little attention. It may be characterized as giving an account of the causes of particular effects; this approach attempts to address the following question: given a particular effect, what

are the various events which might have been its cause? In the contemporary philosophical literature this approach is most notably associated with Mackie's work on insufficient but necessary components of unnecessary but sufficient conditions for an effect (Mackie, 1965). In the epidemiological literature the approach is most closely associated with Rothman's work on sufficient-component causes (Rothman, 1976).

Rothman conceived of a sufficient cause as a minimal set of actions, events or states of nature which together inevitably initiated a set of events resulting in the outcome under consideration. For a particular outcome it is likely that there would be many different sufficient causes, each involving various component causes. Whenever all components of a particular sufficient cause were present, the outcome would inevitably occur; within every sufficient cause, each component would be necessary for that sufficient cause to lead to the outcome. For example, a sufficient cause for some outcome D might consist of the concurrence of conditions A , B and C ; another sufficient cause might be the concurrence of conditions A , F and Q ; and a third sufficient cause might be the concurrence of conditions \bar{Q} and W , where \bar{Q} denotes the complement of Q . These series of conditions, A, B, C and A, F, Q and \bar{Q}, W may each represent different causal mechanisms for the outcome D . When every component of a particular series is present, the outcome D will occur but each component is necessary for the mechanism to be set in motion; thus A, B, C together are sufficient for outcome D but A, B together, without C , is not. Rothman (1976) defined synergism between two causes, A and B say, as the co-participation of A and B in the same sufficient cause. Thus, in the example above, it would be said that A and B exhibit synergism but that F and W do not.

In this section we make formal these notions of sufficient causes. First, we define a sufficient cause and give also a number of related definitions and, secondly, we show in Theorem 1 that for binary variables any counterfactual response pattern in the potential outcomes framework can be replicated by a set of sufficient causes. Throughout this paper, we use the following notation. We let Ω denote the sample space of individuals in the population and we use ω for a particular sample point. An event is a binary variable taking values in $\{0, 1\}$. We denote the complement of an event X by \bar{X} . A conjunction or product of events X_1, \dots, X_n will be written as $X_1 \dots X_n$. The disjunctive or 'or' operator, \vee , is defined by $A \vee B = A + B - AB$, so that $A \vee B = 1$ if and only if either $A = 1$ or $B = 1$.

Consider a potential outcomes framework with s binary factors, X_1, \dots, X_s , which represent hypothetical interventions or causes, and let D denote some binary outcome of interest. Let $D_{x_1 \dots x_s}(\omega)$ denote the counterfactual value of D for individual ω if the causes X_j were set to the value x_j for $j = 1, \dots, s$. We use $D_{x_1 \dots x_s}(\omega)$ and $D_{X_1=x_1, \dots, X_s=x_s}(\omega)$ interchangeably. Note that the potential outcomes framework that we employ in this paper assumes that the counterfactual variables are deterministic, not stochastic. Most of the potential-outcomes literature and the entirety of the sufficient-component cause literature to date assumes a deterministic counterfactual framework, and it is thus the deterministic framework that is the focus of this paper. The relationship between a stochastic

sufficient-component cause model and the stochastic counterfactual framework is a topic of current research. In fact, Theorem 4 below has an analogue for stochastic counterfactuals and stochastic sufficient causes. However, the mathematical and conceptual details required for the stochastic setting lie beyond the scope of this paper. In the deterministic setting there will be 2^s potential outcomes for each individual ω in the population, one potential outcome for each possible value of (X_1, \dots, X_s) . The actual value of D for individual ω will be denoted by $D(\omega)$ and the actual value of X_1, \dots, X_s for individual ω will be denoted by $X_1(\omega), \dots, X_s(\omega)$. Mathematically, it could be that $D_{X_1(\omega)\dots X_s(\omega)}(\omega) \neq D(\omega)$; however, we will require the 'consistency' assumption that $D_{X_1(\omega)\dots X_s(\omega)}(\omega) = D(\omega)$, i.e. that the value of D which would have been observed if X_1, \dots, X_s had been set to what they in fact were is equal to the value of D which was observed. Thus the only potential outcome for individual ω that is observed is the potential outcome $D_{X_1(\omega)\dots X_s(\omega)}(\omega)$, the value of D which would have been observed if X_1, \dots, X_s had been set to what they in fact were.

DEFINITION 1 (Sufficient cause). *A set of binary causes X_1, \dots, X_n for D is said to constitute a sufficient cause for D if for all values x_1, \dots, x_s such that $x_1 \dots x_n = 1$ we have that $D_{x_1 \dots x_s}(\omega) = 1$ for all $\omega \in \Omega$.*

DEFINITION 2 (Minimal sufficient cause). *A set of binary causes X_1, \dots, X_n is said to constitute a minimal sufficient cause for D if X_1, \dots, X_n constitute a sufficient cause for D and no proper subset X_{i_1}, \dots, X_{i_k} of X_1, \dots, X_n also constitutes a sufficient cause for D .*

When a complete set of sufficient causes for some outcome is known, then not only is it the case that the realization of each sufficient cause necessarily entails the outcome, but it is also the case that the presence of the outcome necessarily entails the realization of at least one of the sufficient causes. Such a complete set of sufficient causes will be said to be a determinative set of sufficient causes; when all the sufficient causes of a particular set are needed for the set to be determinative then the set is said to be non-redundant.

DEFINITION 3 (Determinative sufficient causes). *A set of sufficient causes for D , M_1, \dots, M_n , each of which may be some product of binary causes of D , is said to be determinative for D if, for all $\omega \in \Omega$, $D_{x_1 \dots x_s}(\omega) = 1$ if and only if x_1, \dots, x_s are such that $M_1 \vee M_2 \vee \dots \vee M_n = 1$.*

DEFINITION 4 (Non-redundant sufficient causes). *If M_1, \dots, M_n is a determinative set of (minimal) sufficient causes for D such that there is no proper subset M_{i_1}, \dots, M_{i_k} of M_1, \dots, M_n that is also a determinative set of (minimal) sufficient causes for D then M_1, \dots, M_n is said to constitute a non-redundant determinative set of (minimal) sufficient causes for D .*

It will be helpful to distinguish between the concepts of minimality and non-redundancy. Minimality makes references to the components in a particular conjunction, namely that each component is necessary for the conjunction to be

sufficient for the outcome D . Non-redundancy makes reference to a disjunction of conjunctions, that each conjunction is necessary for the disjunction to be determinative. Corresponding to the definition of a sufficient cause is the more philosophical notion of a causal mechanism. A causal mechanism can be conceived of as a set of events or conditions which, if all present, inevitably bring about the outcome under consideration in a particular manner. A causal mechanism thus provides a particular description of how the outcome comes about. We will make reference to the concept of a causal mechanism in some of the discussion of this paper. However, all definitions and theorems are given in terms of sufficient causes for which we have a precise definition. For a sufficient cause to correspond to a particular causal mechanism it is not necessary that the sufficient cause be a minimal sufficient cause nor that it be part of a set of sufficient causes that is non-redundant. This is illustrated in Example 1.

Example 1. Suppose that an individual were exposed to two poisons, X_1 and X_2 , such that, in the absence of X_2 , the poison X_1 would lead to heart failure resulting in death, and, in the absence of X_1 , the poison X_2 would lead to respiratory failure resulting in death, but such that, when X_1 and X_2 were both present, they would interact and lead to a failure of the nervous system once again resulting in death. Here there are three distinct causal mechanisms for death, namely $X_1\bar{X}_2$, \bar{X}_1X_2 and X_1X_2 . Each of these mechanisms is a sufficient cause for death but none of them is minimally sufficient since either X_1 or X_2 alone is sufficient for death.

Although the concepts of minimality of sufficient causes and of non-redundancy are not essential for a sufficient cause to correspond to a causal mechanism, it will be seen in the following section that these concepts are useful in the development of the theory of sufficient cause interactions.

The relationship between the sufficient-component cause framework and the potential outcomes framework has received some attention in the literature. Greenland & Poole (1988) relate the two in the case of two binary causes. Rothman & Greenland (1998), Greenland & Brumback (2002) and Flanders (2006) provide some further discussion. VanderWeele & Robins (2007) relate the sufficient-component cause framework to the directed acyclic graph causal framework and develop theory concerning the graphical representation of sufficient causes on directed acyclic graphs. To develop conditions for sufficient cause interactions we will need only one result concerning the relationship between the sufficient-component cause framework and potential outcomes which is given in Theorem 1. Throughout §3 and for the remainder of §2 we will restrict our attention to the case of a binary outcome and two binary causes. Section 4 indicates how these results generalize. We show in Theorem 1 that, in the case of a binary outcome and two binary variables, given any potential outcomes response pattern it is always possible to construct sufficient causes for the outcome such that the sufficient causes replicate the potential outcome responses.

THEOREM 1. *Suppose that X_1 and X_2 are binary causes of some binary*

outcome D . There exist binary variables $A_0(\omega), \dots, A_8(\omega)$ such that

$$D = A_0 \vee A_1 X_1 \vee A_2 \overline{X_1} \vee A_3 X_2 \vee A_4 \overline{X_2} \vee A_5 X_1 X_2 \vee A_6 \overline{X_1} X_2 \vee A_7 X_1 \overline{X_2} \vee A_8 \overline{X_1} \overline{X_2} \quad (1)$$

and such that

$$D_{x_1 x_2} = A_0 \vee A_1 x_1 \vee A_2 (1 - x_1) \vee A_3 x_2 \vee A_4 (1 - x_2) \vee A_5 x_1 x_2 \vee A_6 (1 - x_1) x_2 \vee A_7 x_1 (1 - x_2) \vee A_8 (1 - x_1) (1 - x_2). \quad (2)$$

Proof. We construct A_0, \dots, A_8 as follows. Let $A_0(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 0$ if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$; let $A_1(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 3$ if $D_{00}(\omega) = 0, D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$; let $A_2(\omega) = A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2, 3$ if $D_{10}(\omega) = 0, D_{00}(\omega) = D_{01}(\omega) = D_{11}(\omega) = 1$; let $A_3(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 3$ if $D_{00}(\omega) = D_{10}(\omega) = 0, D_{01}(\omega) = D_{11}(\omega) = 1$; let $A_1(\omega) = A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1, 4$ if $D_{01}(\omega) = 0, D_{00}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 1$; let $A_1(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 1$ if $D_{00}(\omega) = D_{01}(\omega) = 0, D_{10}(\omega) = D_{11}(\omega) = 1$; let $A_5(\omega) = A_8(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 5, 8$ if $D_{01}(\omega) = D_{10}(\omega) = 0, D_{00}(\omega) = D_{11}(\omega) = 1$; let $A_5(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 5$ if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 0, D_{11}(\omega) = 1$; let $A_2(\omega) = A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2, 4$ if $D_{11}(\omega) = 0, D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 1$; let $A_6(\omega) = A_7(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 6, 7$ if $D_{00}(\omega) = D_{11}(\omega) = 0, D_{01}(\omega) = D_{10}(\omega) = 1$; let $A_2(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 2$ if $D_{10}(\omega) = D_{11}(\omega) = 0, D_{00}(\omega) = D_{01}(\omega) = 1$; let $A_6(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 6$ if $D_{00}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0, D_{01}(\omega) = 1$; let $A_4(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 4$ if $D_{01}(\omega) = D_{11}(\omega) = 0, D_{00}(\omega) = D_{10}(\omega) = 1$; let $A_7(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 7$ if $D_{00}(\omega) = D_{01}(\omega) = D_{11}(\omega) = 0, D_{10}(\omega) = 1$; let $A_8(\omega) = 1$ and $A_i(\omega) = 0$ for $i \neq 8$ if $D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0, D_{00}(\omega) = 1$; let $A_i(\omega) = 0$ for all i if $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = D_{11}(\omega) = 0$. It is then easily verified that (2) holds and (1) then follows by the consistency assumption. \square

Theorem 1 generalizes to the case of a binary outcome and an arbitrary number of binary causes (as shown in an unpublished University of Chicago technical report by the authors). The theorem allows for the construction of variables A_i such that the A_i variables along with X_1 and X_2 and their complements can be used to form a determinative set of sufficient causes for D which replicate a given set of potential outcomes. The conjunctions $A_0, A_1 X_1, \dots, A_8 \overline{X_1} \overline{X_2}$ are sufficient for D and the disjunction of these conjunctions is determinative for D . The variables X_1 and X_2 are causes of D ; the A_i variables are logical constructs and may or may not allow for interpretation; it may not be possible to intervene on these logical constructs. The A_i variables essentially represent unmeasured or unknown factors that complete the particular sufficient cause. Although it may not be possible to intervene on A_i , we will still refer to conjunctions in equation (1) as sufficient causes for D . Theorem 1 and Theorems 2 and 3 below make no assumption as to whether or not these A_i variables are confounding factors or whether or not the causal effects of X_1 and X_2 on D are unconfounded. We will,

however, take up the issue of confounding factors in Theorem 4, which concerns empirical conditions for sufficient cause interactions.

Note that the logical constructs A_i , being functions of the potential outcomes themselves, are not affected by the causes or interventions X_1 and X_2 . If the counterfactual response pattern for every individual in the population is identical, i.e. if the causes X_1 and X_2 completely determine the outcome D , then no additional variables A_i are needed to form a determinative set of sufficient causes for D . A determinative set of sufficient causes for D can be constructed simply from the binary causes X_1 and X_2 and their complements. If for some i $A_i(\omega) = 0$ for all ω then the conjunction in which A_i appears will be suppressed from the disjunction. If for some i $A_i(\omega) = 1$ for all ω then A_i will be suppressed from the conjunction in which it appears but the X_1 and X_2 terms and their complements will appear in the conjunction for the sufficient cause. This is illustrated in Example 2 below. Example 2 also demonstrates that the construction of the A_i variables in the proof of Theorem 1 is not in fact the only possible construction of the A_i variables such that equations (1) and (2) hold; that is to say, a determinative set of sufficient causes will not in general be unique.

Example 2. Consider the example of the two poisons, X_1 and X_2 , presented in Example 1. Suppose now there are two individuals in the population and that individual 2 is immune to poison X_1 and susceptible to poison X_2 , but that individual 1 is susceptible to both poisons. The construction of the A_i variables given in Theorem 1 would give

$$D(\omega) = A_1(\omega)X_1(\omega) \vee X_2(\omega), \quad (3)$$

where $A_1(\omega) = 1$ if $\omega = 1$ and $A_3(\omega) = 1$ for all ω . Note that, since $A_3(\omega) = 1$ for all ω , $A_3(\omega)$ is suppressed from equation (3). We could, however, also define $A_3(\omega) = 1$ for $\omega = 2$; $A_5(\omega) = 1$ for $\omega = 1$; $A_6(\omega) = 1$ for all ω ; $A_7(\omega) = 1$ for $\omega = 1$; and $A_i(\omega) = 0$ for all ω for all other i . Equations (1) and (2) would again hold and we could write

$$D(\omega) = A_3(\omega)X_2(\omega) \vee A_5(\omega)X_1(\omega)X_2(\omega) \vee \overline{X_1}(\omega)X_2(\omega) \vee A_7(\omega)X_1(\omega)\overline{X_2}(\omega). \quad (4)$$

Other constructions of the A_i variables are also possible. However, in this case, it is the construction given in (4) rather than that given in (3) which arguably corresponds better to the actual causal mechanisms. The sufficient cause $A_5X_1X_2$ may be interpreted as death by failure of the nervous system; the sufficient cause $A_7X_1\overline{X_2}$ may be interpreted as death by heart failure; and the sufficient causes A_3X_2 and $\overline{X_1}X_2$ may be interpreted as death by respiratory failure. No similar interpretation holds for the construction given for (3).

As noted above, some knowledge of the subject matter in question will be necessary in order to determine which of several possible constructions of the A_i variables corresponds best to the actual causal mechanisms for the outcome D . However, any set of binary variables A_i such that the disjunction $A_0 \vee A_1X_1 \vee A_2\overline{X_1}$

$\bigvee A_3 X_2 \bigvee A_4 \overline{X_2} \bigvee A_5 X_1 X_2 \bigvee A_6 \overline{X_1} X_2 \bigvee A_7 X_1 \overline{X_2} \bigvee A_8 \overline{X_1} \overline{X_2}$ replicates the potential outcome response patterns for the entire population we will call a sufficient cause representation for D .

DEFINITION 5 (Sufficient cause representation). *Let X_1 and X_2 be binary causes of some binary outcome D , we will say that*

$$A_0 \bigvee A_1 X_1 \bigvee A_2 \overline{X_1} \bigvee A_3 X_2 \bigvee A_4 \overline{X_2} \bigvee A_5 X_1 X_2 \bigvee A_6 \overline{X_1} X_2 \bigvee A_7 X_1 \overline{X_2} \bigvee A_8 \overline{X_1} \overline{X_2} \quad (5)$$

constitutes a sufficient cause representation for D for any set of binary variables $A_0(\omega), \dots, A_8(\omega)$ such that equation (2) holds.

For any sufficient cause representation, each conjunction within the disjunction (5) is a sufficient cause for the outcome D and the collection of conjunctions constitutes a determinative set of sufficient causes for D . If the conjunctions in a particular sufficient cause representation are minimal sufficient causes then we will refer to the representation as a minimal sufficient cause representation. If the conjunctions in a particular sufficient cause representation are non-redundant then we will refer to the representation as a non-redundant sufficient cause representation. With these definitions in place we can now derive conditions which imply the existence of sufficient cause interactions.

3. SUFFICIENT CAUSE INTERACTIONS

In this section we define and develop conditions for testing for the presence of sufficient cause interactions. We begin formally by defining the concepts of a sufficient cause interaction and a minimal sufficient cause interaction and showing how they are related.

DEFINITION 6 (Minimal sufficient cause interaction). *Let F_1 be either X_1 or its complement and let F_2 be either X_2 or its complement. Then $F_1 F_2$ is said to exhibit a minimal sufficient cause interaction if in every non-redundant minimal sufficient cause representation for D there exists within the representation a sufficient cause which contains $F_1 F_2$ within its conjunction.*

In Example 2, it can be verified that (3) constitutes a minimal sufficient cause representation for D . Since the term $X_1 X_2$ does not appear in this disjunction, $X_1 X_2$ does not exhibit a minimal sufficient cause interaction.

Corresponding to the definition of a minimal sufficient cause interaction is that of a sufficient cause interaction, which makes reference to all sufficient cause representations and not just to non-redundant minimal sufficient cause representations.

DEFINITION 7 (Sufficient cause interaction). *Let F_1 be either X_1 or its complement and let F_2 be either X_2 or its complement. Then $F_1 F_2$ is said to exhibit a sufficient cause interaction, or to be irreducible, if in every sufficient cause representation for D there exists within the representation a sufficient cause which contains $F_1 F_2$ within its conjunction.*

Theorem 2 demonstrates that the concepts of a sufficient cause interaction (irreducibility) and a minimal sufficient cause interaction are equivalent. The theorem will be used in our derivation of counterfactual and empirical conditions for sufficient cause interactions.

THEOREM 2. *The conjunction F_1F_2 is irreducible if and only if F_1F_2 exhibits a minimal sufficient cause interaction.*

Proof. If F_1F_2 is irreducible then within any sufficient cause representation there exists some sufficient cause which contains within its conjunction F_1F_2 and so it immediately follows that in every non-redundant minimal sufficient cause representation for D there will exist within the representation a sufficient cause which contains F_1F_2 in its conjunction. If F_1F_2 is not irreducible then there exists some representation such that no sufficient cause within the representation contains F_1F_2 within its conjunction. This representation can be made into a non-redundant minimal sufficient cause representation by iteratively discarding the components of each conjunction which are not necessary for the conjunction to be sufficient for D and then iteratively discarding any redundant minimal sufficient causes. Clearly no sufficient cause of this resulting non-redundant minimal sufficient causation representation will contain F_1F_2 within its conjunction. \square

We will say that there is synergism between the effects of F_1 and F_2 on D if there exists a sufficient cause for D with F_1F_2 in its conjunction which represents a particular causal mechanism for D . Example 2 above suggests that some knowledge of the causal mechanisms beyond that which is available by a complete knowledge of the counterfactual outcomes may be required to determine whether or not synergism between F_1 and F_2 is present. In Example 2, it is not possible to distinguish merely from the counterfactual outcomes whether $A_1X_1 \vee X_2$ or $A_3X_2 \vee A_5X_1X_2 \vee \overline{X_1}X_2 \vee A_7X_1\overline{X_2}$ or some other sufficient cause representation constitutes the proper description of the causal mechanisms for D . It is thus not possible to determine in this example from the counterfactual outcomes alone whether or not there is synergism between X_1 and X_2 . The presence of synergism will sometimes be unidentified even when the counterfactual outcomes for all individuals are known. As was the case with the concept of a causal mechanism, statements about synergism will in general require some knowledge of the subject matter in question. In Example 2, our knowledge of the causal mechanisms by which poisons X_1 and X_2 operate allow us to determine that the sufficient cause representation $A_3X_2 \vee A_5X_1X_2 \vee \overline{X_1}X_2 \vee A_7X_1\overline{X_2}$ corresponds to the actual causal mechanisms for death and thus that synergism is present. Note that, in Example 2, synergism between X_1 and X_2 is present even though X_1X_2 does not exhibit a sufficient cause interaction. The conjunction X_1X_2 does not exhibit a sufficient cause interaction because $A_1X_1 \vee X_2$ constitutes a sufficient cause representation for D and the term X_1X_2 is not present in any of the conjunctions in this representation.

Although the presence of synergism is sometimes unidentified from the complete set of counterfactual outcomes as in Example 2, it is not always unidentified. If the conjunction F_1F_2 is irreducible then within every sufficient cause representation for D there exists some sufficient cause which contains F_1F_2 within its conjunction, and so there must be some causal mechanism for which F_1F_2 are required; synergism must be present. The class of conjunctions which are irreducible, or equivalently the components of which exhibit a minimal sufficient cause interaction, are the class for which synergism must be present. The results we give below will be stated in terms of the well-defined concept of a sufficient cause interaction. However, it is the more philosophical notions of synergism and causal mechanism that provide much of the motivation for these results. We will, in the interpretation of our results, assume that there always exists some set of true causal mechanisms which forms a determinative set of sufficient causes for the outcome. Theorem 3 relates the class of irreducible sufficient cause representations explicitly to counterfactual outcomes. Theorem 4 demonstrates that in certain cases one can conclude from data that a particular conjunction is irreducible and thus that synergism must be present.

THEOREM 3. *The conjunction X_1X_2 exhibits a sufficient cause interaction if and only if there exists $\omega \in \Omega$ such that $D_{11}(\omega) = 1$ and $D_{10}(\omega) = D_{01}(\omega) = 0$.*

Proof. Suppose that X_1X_2 do not exhibit a minimal sufficient cause interaction. Then there exists a non-redundant minimal sufficient cause representation such that equations (1) and (2) hold and such that $A_i(\omega) = 0$ for all ω for $i \in \{5, 6, 7, 8\}$. For any ω such that $D_{10}(\omega) = D_{01}(\omega) = 0$ we must have that $A_0(\omega) = A_1(\omega) = A_3(\omega) = 0$ and thus that $D_{11}(\omega) = A_2(\omega)(1-1) \vee A_4(\omega)(1-1) = 0$. Consequently there can be no ω such that $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$. We now prove the converse. Suppose that X_1X_2 do exhibit a minimal sufficient cause interaction. Then within every sufficient cause representation for D there exists within the representation a sufficient cause which contains X_1X_2 within its conjunction, and so, in particular in the sufficient cause representation constructed in Theorem 1, there exists an ω such that $A_5(\omega) \neq 0$. However, by construction $A_5(\omega) \neq 0$ if and only if either (1) $D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{00}(\omega) = D_{11}(\omega) = 1$ or (2) $D_{00}(\omega) = D_{01}(\omega) = D_{10}(\omega) = 0$ and $D_{11}(\omega) = 1$. In either case, $D_{10}(\omega) = D_{01}(\omega) = 0$ and $D_{11}(\omega) = 1$. This completes the proof. \square

The condition provided in Theorem 3 has obvious analogues if one or both of X_1 and X_2 are replaced with their complements. Theorem 3 suggests a very natural empirical condition for detecting the presence of a sufficient cause interaction; this is given in Theorem 4. Some discussion with regard to constructing statistical tests related to this condition is given below. In §5, the condition stated in Theorem 4 is related explicitly to statistical tests arising from generalized linear models.

THEOREM 4. *Let C be any set of variables which suffices to control for the confounding of the causal effects of X_1 and X_2 on D , i.e. such that $D_{x_1x_2} \perp\!\!\!\perp \{X_1, X_2\} | C$.*

Then, if for any value c of C we have that

$$\begin{aligned} & E(D|X_1 = 1, X_2 = 1, C = c) - E(D|X_1 = 0, X_2 = 1, C = c) \\ & - E(D|X_1 = 1, X_2 = 0, C = c) > 0, \end{aligned} \quad (6)$$

it follows that X_1X_2 exhibit a sufficient cause interaction.

Proof. We prove the contrapositive. Suppose that X_1X_2 does not exhibit a sufficient cause interaction between then by Theorem 3 it would follow that there is no $\omega \in \Omega$ such that $D_{11}(\omega) = 1$ and $D_{10}(\omega) = D_{01}(\omega) = 0$. From this it follows that for all $\omega \in \Omega$ we have $D_{11}(\omega) - D_{10}(\omega) - D_{01}(\omega) \leq 0$ and so $E(D_{11} - D_{10} - D_{01}) \leq 0$. Since $D_{x_1x_2} \coprod\coprod\{X_1, X_2\}|C$ we have that

$$\begin{aligned} & E(D|X_1 = 1, X_2 = 1, C = c) - E(D|X_1 = 0, X_2 = 1, C = c) \\ & - E(D|X_1 = 1, X_2 = 0, C = c) \\ = & E(D_{11}|X_1 = 1, X_2 = 1, C = c) - E(D_{01}|X_1 = 0, X_2 = 1, C = c) \\ & - E(D_{10}|X_1 = 1, X_2 = 0, C = c) \\ = & E(D_{11}|C = c) - E(D_{01}|C = c) - E(D_{10}|C = c) \\ = & E\{D_{11}(\omega) - D_{10}(\omega) - D_{01}(\omega)\} \leq 0. \end{aligned}$$

This completes the proof. \square

As with Theorem 3, the condition provided in Theorem 4 has obvious analogues if one or both of X_1 and X_2 are replaced with their complements. The conditions given in Theorems 3 and 4 are novel in that, unlike previous conditions given in the literature, they make no monotonicity assumption about the relationship between the outcome D and the causes, X_1 and X_2 .

Previous work on identifying synergism focused on monotonicity assumptions. If $D_{x_1x_2}(\omega)$ is nondecreasing in x_1 and x_2 for all ω , we will say that X_1 and X_2 have monotonic effects on D . It can be shown that, if X_1 and X_2 have monotonic effects on D and if C is set of variables which suffices to control for the confounding of the causal effects of X_1 and X_2 on D , i.e. such that $D_{x_1x_2} \coprod\coprod\{X_1, X_2\}|C$, then, if for any value c of C we have that

$$\begin{aligned} & E(D|X_1 = 1, X_2 = 1, C = c) - E(D|X_1 = 0, X_2 = 1, C = c) \\ & - E(D|X_1 = 1, X_2 = 0, C = c) + E(D|X_1 = 0, X_2 = 0, C = c) > 0, \end{aligned} \quad (7)$$

it follows that X_1X_2 exhibit a sufficient cause interaction. In the context of no confounding factors a result equivalent to this, without the formalization of the concept of a sufficient cause interaction, is stated explicitly and proved by Rothman & Greenland (1998); it is also anticipated elsewhere (Koopman, 1981; Darroch & Borkent, 1994). It is instructive to compare condition (7) with the condition derived in Theorem 4 without the assumption of monotonicity. The assumption of monotonicity effectively allows for the addition of the term $E(D|X_1 = 0, X_2 = 0, C = c)$; condition (7) is thus clearly weaker than condition (6) and consequently could be used in the construction of more powerful tests for

the presence of a sufficient cause interaction. Condition (6) is novel in that it requires no assumption about monotonicity. It will be seen in §5 that, unlike condition (6), condition (7) is quite intuitive in that it obviously corresponds to departures from additivity.

Conditions (6) and (7) can be used to test empirically for the presence of sufficient cause interactions and thus for synergism. If the set of confounding variables C consists of a small number of binary or categorical variables then it may be possible to use t -test-like test statistics 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 $\text{pr}(D = 1|X_1, \dots, X_m, C)$ using a binomial or Poisson regression model with a linear link (Greenland, 1991; Wacholder, 1986; Zou, 2004; Greenland, 2004; Spiegelman & Hertzmark, 2005). Unfortunately, when continuous covariates are included in Bernoulli regressions with linear link, the convergence properties of maximum likelihood estimators are generally poor (Wacholder, 1986). For case-control studies it will also be necessary to use an adapted set of modelling techniques (Wild, 1991; Wacholder, 1996; Greenland, 2004). We are currently working on doubly robust semiparametric tests for hypotheses (6) and (7) where C is multivariate with possibly continuous confounding variables.

4. GENERALIZATIONS AND EXTENSIONS

The definitions of a sufficient cause representation and sufficient cause interactions and minimal sufficient cause interactions generalize to cases in which n binary causes of some binary outcome D are being considered. Theorems 1-4 also generalize to the case of sufficient cause representation for n binary causes and n -way sufficient cause interactions. Full discussion of these generalizations is beyond the scope of the present paper; a detailed exposition is available from the authors upon request, in the form of the technical report mentioned in §2. However, for the purposes of comparing sufficient cause interactions and statistical interactions in the following section it will be helpful at least to indicate how Theorems 3 and 4 generalize to the case of three-way sufficient cause interactions.

It can be shown that for three binary causes, X_1 , X_2 and X_3 , of some binary outcome D , if there exists an individual ω for whom $D_{111}(\omega) = 1$ and $D_{011}(\omega) = D_{101}(\omega) = D_{110}(\omega) = 0$ then there exists a three-way sufficient cause interaction between X_1 , X_2 and X_3 . From this it follows that, if C is any set of variables which suffices to control for the confounding of the causal effects of X_1 , X_2 and X_3 on D , i.e. such that $D_{x_1 \dots x_n} \perp\!\!\!\perp \{X_1, X_2, X_3\} | C$, then if for any value c of C

$$\begin{aligned} & E(D|X_1 = 1, X_2 = 1, X_3 = 1, C = c) - E(D|X_1 = 0, X_2 = 1, X_3 = 1, C = c) \\ & - E(D|X_1 = 1, X_2 = 0, X_3 = 1, C = c) - E(D|X_1 = 1, X_2 = 1, X_3 = 0, C = c) > 0, \end{aligned} \tag{8}$$

there exists a three-way sufficient cause interaction between X_1 , X_2 and X_3 . Like condition (6) in Theorem 4, condition (8) can be used to construct empirical statistical tests for the presence of sufficient cause interactions and thus for synergism.

Condition (8) also has obvious analogues if one or more of X_1 , X_2 and X_3 are replaced with their complements.

Note that condition (6) concerns two-way interactions in a two-cause sufficient-component cause framework and that condition (8) concerns three-way interactions in a three-cause sufficient-component cause framework. Additional subtleties arise if, for example, one considers two-way interactions in a three-cause sufficient-component cause framework. Essentially, to test for a two-way interaction in a three-cause sufficient-component cause framework, it is necessary to assume that some of the causes are not effects of the others. For example, consider three binary causes, X_1 , X_2 and X_3 , of some binary variable D and suppose that we were interested in testing for the presence of a sufficient cause interaction between X_1 and X_2 . To use Theorems 3 and 4 to test for the presence of a two-way sufficient cause interaction it would also be necessary to assume that neither X_1 nor X_2 was a cause of X_3 . Otherwise X_3 might serve as a proxy for the joint presence of X_1 and X_2 as would be the case, for example, if X_3 were 1 whenever X_1 and X_2 were both set to 1. When the causes X_1 , X_2 and X_3 are such that none of these variables is a cause of another, these difficulties do not arise; see the aforementioned technical report for further details.

Conditions for a sufficient cause interaction under assumptions of monotonicity can also be generalized to n -way sufficient cause interactions. The conditions are, however, somewhat more complicated. For example, in the case of three binary variables, X_1 , X_2 and X_3 , which all have monotonic effects on the outcome D , if any of the following three conditions are met, then a three-way sufficient cause interaction must be present:

$$\begin{aligned} p_{111c} - p_{110c} - p_{101c} - p_{011c} + p_{100c} + p_{010c} &> 0 \\ p_{111c} - p_{110c} - p_{101c} - p_{011c} + p_{100c} + p_{001c} &> 0 \\ p_{111c} - p_{110c} - p_{101c} - p_{011c} + p_{010c} + p_{001c} &> 0, \end{aligned} \tag{9}$$

where $p_{x_1x_2x_3c} = E(D|X_1 = x_1, X_2 = x_2, X_3 = x_3, C = c)$. Further detail is given in the aforementioned technical report.

5. SUFFICIENT CAUSE INTERACTIONS AND INTERACTION TERMS IN STATISTICAL MODELS

The results given above provide conditions which can be empirically tested to draw inferences about the presence of sufficient cause interactions. If a sufficient cause interaction between X_1 and X_2 is present, then in any sufficient cause representation for the outcome there must exist a sufficient cause in which the conjunction X_1X_2 is present; the causal mechanisms for the outcome must be such that X_1 and X_2 both participate in the same causal mechanism. Conditions (6)-(9) above are given in terms of differences between various probabilities. To understand these conditions better we will consider saturated Bernoulli regression models with linear links. For simplicity we will assume that the causal effects of $\{X_1, X_2\}$ or $\{X_1, X_2, X_3\}$ are unconfounded. However, the substance of the

remarks below are not altered in the case of one or more binary or categorical confounding variables. We will compare, within a regression framework, the tests arising from conditions (6)-(9) with standard tests for statistical interactions.

We will begin with the case of two-way interactions. Consider a saturated Bernoulli regression model,

$$\text{pr}(D = 1|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2.$$

We will use $p_{x_1x_2}$ as a shorthand for $\text{pr}(D = 1|X_1 = x_1, X_2 = x_2)$. In this statistical model, one would test for a statistical interaction by testing the hypothesis $\beta_3 = 0$. We will consider first the case of monotonic effects. If X_1 and X_2 have monotonic effects on D , condition (7) states that if

$$p_{11} - p_{10} - p_{01} + p_{00} > 0$$

then there exists a sufficient cause interaction between X_1 and X_2 . We may rewrite this condition as

$$p_{11} - p_{10} - p_{01} + p_{00} = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) - (\beta_0 + \beta_2) + \beta_0 = \beta_3 > 0.$$

In the case of monotonic effects, if the statistical interaction term β_3 is positive then a sufficient cause interaction is necessarily present between X_1 and X_2 . If it cannot be assumed that X_1 and X_2 have monotonic effects on D we may apply Theorem 4, which in this case states that a sufficient cause interaction between X_1 and X_2 will be present if

$$p_{11} - p_{10} - p_{01} > 0,$$

which can be rewritten as

$$\beta_3 > \beta_0.$$

We see then that the tests for statistical interaction only correspond to tests for sufficient cause interactions in the case of monotonic effects, not in general. Furthermore, even under monotonic effects, a statistical interaction only implies a sufficient cause interaction if the interaction coefficient β_3 is positive; if β_3 is non-zero but negative, we cannot draw conclusions about the presence of a sufficient cause interaction.

We now consider the case of a three-way sufficient cause interaction. The saturated Bernoulli regression with three binary variables and a linear link can be written as

$$\begin{aligned} \text{pr}(D = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3) = & \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2 \\ & + \beta_5x_1x_3 + \beta_6x_2x_3 + \beta_7x_1x_2x_3. \end{aligned}$$

Once again we will use the shorthand $p_{x_1x_2x_3} = \text{pr}(D = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3)$. The presence of a three-way statistical interaction would be assessed by testing the hypothesis $\beta_7 = 0$. Under the assumption that X_1 , X_2 and X_3 have monotonic effects on D , condition (9) states that X_1 , X_2 and X_3 exhibit a sufficient

cause interaction if any of the following three conditions hold:

$$\begin{aligned} p_{111} - p_{110} - p_{101} - p_{011} + p_{100} + p_{010} &> 0 \\ p_{111} - p_{110} - p_{101} - p_{011} + p_{100} + p_{001} &> 0 \\ p_{111} - p_{110} - p_{101} - p_{011} + p_{010} + p_{001} &> 0. \end{aligned}$$

These three conditions can be rewritten in terms of the regression coefficients as follows:

$$\begin{aligned} \beta_7 &> \beta_3 \\ \beta_7 &> \beta_2 \\ \beta_7 &> \beta_1. \end{aligned}$$

If it cannot be assumed that X_1 , X_2 and X_3 have monotonic effects on D we may use condition (8), which in this case states that X_1 , X_2 and X_3 exhibit a sufficient cause interaction if

$$p_{111} - p_{110} - p_{101} - p_{011} > 0,$$

which is equivalent to

$$\beta_7 > 2\beta_0 + \beta_1 + \beta_2 + \beta_3.$$

In the case of three-way sufficient cause interactions we thus see that neither the tests for a sufficient cause interaction under the assumption of monotonic effects nor the tests without the assumption of monotonic effects are equivalent to the standard hypothesis test for a three-way statistical interaction.

6. CONCLUDING REMARKS

In this paper we have derived various conditions which, if met, necessarily entail that two binary causes participate in a single causal mechanism. Unlike previous conditions in the literature, the conditions we have derived make no assumption about monotonicity. Under the assumption of no unmeasured confounding variables, these conditions can, with data, be empirically tested. We have shown that interaction terms in standard statistical models do not capture the form of interdependence which we have characterized as a sufficient cause interaction. This work may be of special interest to statistical geneticists in identifying gene-gene and gene-environment interactions. The gene-gene and gene-environment interdependence that is ultimately of interest to the geneticist will often not be that of association but of mechanism. The tests we have derived are concerned with mechanistic interaction.

A couple of limitations of the present work are worth noting. First, the tests derived here are applicable only when the outcome and the causes under consideration are all binary. If causation is fundamentally a phenomenon concerning events (Lewis, 1973a; Davidson, 1980; Lewis, 1986) then the restriction to binary causes is not, in principal, a limitation. However, in practice, precluding continuous variables will limit the settings in which the methods can be applied. A

second limitation of this work concerns the cases in which synergism is present but a sufficient cause interaction is not. As noted in the text the conditions that entail a sufficient cause interaction are sufficient but not necessary for two causes to participate in the same causal mechanism, i.e. for synergism to be present. Synergism can be present even if conditions (6)-(9) do not hold. Such instances of synergism cannot be identified from data.

We intend to extend the present work by deriving semiparametric tests of hypotheses (6)-(9) and by providing results that relate the stochastic counterfactual framework to stochastic sufficient-component cause models.

REFERENCES

- DARROCH, J.N. & BORKENT, M. (1994). Synergism, attributable risk and interaction for two binary exposure factors. *Biometrika* **81**, 259-70.
- DAVIDSON, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.
- FLANDERS, D. (2006). Sufficient-component cause and potential outcome models. *Eur. J. Epidemiol.* **21**, 847-53.
- GREENLAND, S. (1991). Estimating standardized parameters from generalized linear models. *Statist. Med.* **10**, 1069-74.
- 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**, 301-5.
- GREENLAND, S. & BRUMBACK, B. (2002). An overview of relations among causal modelling methods. *Int. J. Epidemiol.* **31**, 1030-37.
- GREENLAND, S. & POOLE, C. (1988). Invariants and noninvariants in the concept of interdependent effects. *Scand. J. Work Envir. Health* **14**, 125-9.
- HUME, D. (1748). *An Enquiry Concerning Human Understanding*. Reprinted, 1958, LaSalle, IL: Open Court Press.
- KOOPMAN, J.S. (1981). Interaction between discrete causes. *Am. J. Epidemiol.* **113**, 716-24.
- LEWIS, D. (1973a). Causation. *J. Phil.* **70**, 556-67.
- LEWIS, D. (1973b). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- LEWIS, D. (1986). Events. In *Philosophical Papers, Vol. 2*. New York: Oxford University Press.
- MANTEL, N., BROWN, C. & BYAR, D.P. (1977). Tests for homogeneity of effect in an epidemiologic investigation. *Am. J. Epidemiol.* **106**, 125-9.
- NEYMAN, J. (1923). Sur les applications de la thar des probabilités aux expériences Agaricales: Essay des principe. Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, Trans.) in *Statist. Sci.* **5**, 463-72.
- ROBINS, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Math. Mod.* **7**, 1393-512.
- ROBINS, J.M. (1987). Addendum to a new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Comp. Math. Applic.* **14**, 923-45.

- ROTHMAN, K.J. (1976). Causes. *Am. J. Epidemiol.* **104**, 587-92.
- ROTHMAN, K.J. & GREENLAND, S. (1998). *Modern Epidemiology*. Philadelphia: Lippincott-Raven.
- RUBIN, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688-701.
- RUBIN, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6**, 34-58.
- SPIEGELMAN, D. & HERTZMARK, E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. *Am. J. Epidemiol.* **162**, 199-200.
- VANDERWEELE, T.J. & ROBINS, J.M. (2007). Directed acyclic graphs, sufficient causes and the properties of conditioning on a common effect. *Am. J. Epidemiol.*, in press.
- WACHOLDER, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *Am. J. Epidemiol.* **123**, 174-84.
- WACHOLDER, S. (1996). The case-control study as data missing by design: Estimating risk differences. *Epidemiology.* **7**, 145-50.
- WILD, C.J. (1991). Fitting prospective regression models to case-control data. *Biometrika* **78**, 705-17.
- ZOU, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *Am. J. Epidemiol.* **159**, 702-6.