Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2006 *Paper* 13

A Theory of Sufficient Cause Interactions

Tyler J. VanderWeele* James M. Robins†

*University of Chicago, tvanderw@hsph.harvard.edu
†Harvard School of Public Health, robins@hsph.harvard.edu

A Theory of Sufficient Cause Interactions

Tyler J. VanderWeele and James M. Robins

Abstract

Sufficient-component causes are discussed within the potential outcome framework so as to formalize notions of sufficient causes, synergism and sufficient cause interactions. Doing so allows for the derivation of counterfactual conditions and statistical tests for detecting the presence of sufficient cause interactions. Under the assumption of monotonic effects, more powerful statistical tests for sufficient cause interactions can be derived. The statistical tests derived for sufficient cause interactions are compared with and contrasted to interaction terms in standard statistical models.

1. Introduction

It is often of interest to researchers whether two variables in some way interact in the effect they produce on a particular outcome. Interaction terms in statistical models are frequently used to assess whether effects are interdependent. However, whether two variables have a statistical interaction may depend on which statistical model is being used (Mantel et al., 1977). Two variables that have an interaction under one statistical model and may not have an interaction under a different statistical model. 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 sufficient-component cause 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 biologic interactions or synergism. In this paper we develop a theory of sufficient cause interactions and derive various conditions to statistically test for the presence of fully general n-way sufficient cause interactions.

In section 2, we give an overview of the sufficient-component cause framework and demonstrate that in the case of a binary outcome and an arbitrary number of binary causes, given any potential outcome response pattern it is always possible to construct sufficient causes for the outcome such that the sufficient causes replicate the potential outcome responses. In section 3, we formally define sufficient-cause interactions and show that there exist certain conditions which, if met, necessarily entail that any sufficient cause representation for the outcome must have an interaction within the sufficient-component cause framework. In section 4, we develop weaker conditions in the setting where the direction of the effect of the causes on the outcome is known. In section 5 we relate the conditions derived in sections 3 and 4 to interaction terms in statistical models. Section 6 concludes with some closing remarks.

2. Sufficient Causes

Two broad conceptualizations of causality can be discerned in the literature, both within philosophy and within statistics and epidemiology. The first conceptualization of causality 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 outcome framework has been used extensively in statistics both in the development of theory and in application. In contrast, the second conceptualization of causality has received comparatively little attention. This second conceptualization may be characterized as giving an account of the causes of particular effects; this approach attempts to address the 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 (INUS conditions) for an effect (Mackie, 1965). In the epidemiologic literature this 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 there would likely be many different sufficient causes. Each sufficient cause involved various component causes. Whenever all components of a particular sufficient cause were present, the outcome would inevitably occur; within every sufficient cause, each component would be necessary for that sufficient cause to lead to the outcome. 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 third sufficient cause might be the concurrence of conditions $\overline{Q}$ and W where $\overline{Q}$ denotes the complement of Q. These series of conditions, A, B, C and A, F, Q and $\overline{Q}, W$ 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 will define a sufficient cause and give also a number of related definitions and second we will 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 will use the following notation. An event is a binary variable taking values in $\{0, 1\}$. The complement of some event X we will denote by $\overline{X}$. A conjunction or product of events $X_1, ..., X_n$ will be written as $X_1...X_n$. The disjunctive or OR operator, $\bigvee$, is defined by $A \bigvee B = A + B - AB$ so that $A \bigvee B = 1$ if and only if either $A = 1$ or $B = 1$. We will use the notation $1(A)$ to denote the indicator function for condition A.

Consider a potential outcomes framework with s binary factors, $X_1, ..., X_s$, which represent hypothetical interventions or causes and let D denote some binary outcome of interest. Let $D_{x_1...x_s}(\omega)$ to denote the counterfactual value of D for individual ω if the causes X_j were set to the value x_j for $j = 1, ..., s$. We will use $D_{x_1...x_s}(\omega)$ and $D_{X_1=x_1,...,X_s=x_s}(\omega)$ interchangeably. In this setting there will be 2^s potential outcomes for each individual ω in the population, one potential outcome for each possible value of $(X_1, ..., X_s)$. The actual value of D for individual ω will be denoted by $D(\omega)$ and the actual value of $X_1, ..., X_s$ for individual ω will be denoted by $X_1(\omega), ..., X_s(\omega)$. Mathematically, it could be that $D_{X_1(\omega)...X_s(\omega)}(\omega) \neq D(\omega)$; however we will require the "consistency" assumption that $D_{X_1(\omega)...X_s(\omega)}(\omega) = D(\omega)$ i.e. that the value of D which would have been observed if $X_1, ..., X_s$ had been set to what they in fact were is equal to the value of D which was in fact observed. Thus the only potential outcome for individual ω that is observed is the potential outcome $D_{X_1(\omega)...X_s(\omega)}(\omega)$, the value of D which would have been observed if $X_1, ..., X_s$ had been set to what they in fact were. All of the potential outcomes for an individual ω can be listed in a vector with 2^s components and this vector we will denote by $\mathcal{D}(\omega)$.

We now begin with the definitions of a sufficient cause and a minimal sufficient cause for some subset $X_1, ..., X_n$ of the causes $X_1, ..., X_s$.

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

Definition 2 (Minimal Sufficient Cause). A set of binary causes $X_1, ..., X_n$ is said to constitute a minimal sufficient cause for D if $X_1, ..., X_n$ constitute a sufficient cause for D and no proper subset $X_{i_1}, ..., X_{i_k}$ of $X_1, ..., X_n$ 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 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, $M_1, ..., M_n$, each of which may be some product of binary causes of D, is said to be determinative for D if for all $\omega \in \Omega$, $D_{x_1...x_s}(\omega) = 1$ if and only if $x_1, ..., x_s$ are such that $M_1 \bigvee M_2 \bigvee ... \bigvee M_n = 1$.

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

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 that 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: $X_1\overline{X_2}$, $\overline{X_1}X_2$, 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 relation between the sufficient-component cause framework and the potential outcomes framework has received some attention in the literature. Greenland and Poole relate the two (1988) in the case of two binary causes. Rothman and Greenland (1998), Greenland and Brumback (2002) and Flanders (2006) provide some further discussion. VanderWeele and Robins (2006a) 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. For the development of a theory of sufficient cause interactions we will need only one result concerning the relation between the sufficient-component cause framework and potential outcomes. We show in Theorem 1 that in the case of a binary outcome and an arbitrary number of binary causes, given any potential outcome 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 $X_1, ..., X_s$ are binary causes of some binary outcome D. Let Ω be the sample space of the individuals in the population and let $D_{x_1...x_s}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_j were set to x_j. For each possible conjunction $G_i = F_1^i...F_{n_i}^i$, where each F_k^i is either a is either a member of the set $\{X_1, ..., X_s\}$ or is the complement of such a member, there exists a binary variable $A_i(\omega)$ which are functions of the potential outcome vector $\mathcal{D}(\omega)$ such that $D(\omega) = \bigvee_i A_i(\omega)F_1^i(\omega)...F_{n_i}^i(\omega)$ and such that $D_{x_1...x_s}(\omega) = \bigvee_i A_i(\omega)g_i(x_1, ..., x_s)$ where $g_i(x_1, ..., x_s) = 1$ if $F_1^i...F_{n_i}^i = 1$ when $(X_1, ..., X_s) = (x_1, ..., x_s)$ and 0 otherwise.

Theorem 1 allows for the construction of variables A_i such that the A_i variables along with $X_1, ..., X_s$ 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 $A_iF_1^i...F_{n_i}^i$ are sufficient for D and the disjunction of these conjunctions is determinative for D. Each F_k^i in these conjunctions is a cause of D since each F_k^i is either a member of the set $\{X_1, ..., X_s\}$ or is the complement of such a member. The variable A_i is a logical construct and may or may not allow for interpretation; it may not be possible to intervene on this logical construct A_i. Although it may not be possible to intervene on A_i, we will still refer to conjunctions of the form $A_i F_1^i ... F_{n_i}^i$ as sufficient causes for D. Note that the logical constructs A_i, being functions of the potential outcomes themselves, are not affected by any of the causes or interventions $X_1, ..., X_s$. If the counterfactual response pattern for every individual in the population is identical, i.e. if the causes $X_1, ..., X_s$ 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 set of binary causes $X_1, ..., X_s$ and their complements.

The construction in the proof of Theorem 1 of the binary variables A_i is not in fact the only possible construction such that the disjunction $\bigvee_i A_i F_1^i ... F_{n_i}^i$ replicates the potential outcome response patterns $\mathcal{D}(\omega)$. A determinative set of sufficient causes will not in general be unique. For example, in the case of one binary cause, X_1, if $D = A_0 \bigvee A_1 X$ then it is also the case that $D = B_0 \bigvee B_1 X$ where $B_0 = A_0$ and $B_1 = \overline{A_0} A_1$. Any set of binary variables $A_i(\omega)$ constructed from the potential outcomes $\mathcal{D}(\omega)$ such that the disjunction $\bigvee_i A_i F_1^i ... F_{n_i}^i$ replicates the potential outcome response patterns for the entire population we will call a sufficient cause representation for D.

Definition 5 (Sufficient Cause Representation). Suppose that $X_1, ..., X_s$ are binary causes of some binary outcome D. Let Ω be the sample space of the individuals in the population and let $D_{x_1 ... x_s}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_j were set to x_j. For each possible conjunction $G_i = F_1^i ... F_{n_i}^i$, where each F_k^i is either a is either a member of the set $\{X_1, ..., X_s\}$ or is the complement of such a member let $A_i(\omega)$ be any binary variable constructed from the potential outcomes $\mathcal{D}(\omega)$. If $D_{x_1 ... x_s}(\omega) = \bigvee_i A_i(\omega) g_i(x_1, ..., x_s)$ where $g_i(x_1, ..., x_s) = 1$ if $F_1^i ... F_{n_i}^i = 1$ when $(X_1, ..., X_s) = (x_1, ..., x_s)$ and 0 otherwise then the disjunction $\bigvee_i A_i F_1^i ... F_{n_i}^i$ is said to be a sufficient cause representation for D.

For any sufficient cause representation, $\bigvee_i A_i F_1^i ... F_{n_i}^i$, each conjunction $A_i F_1^i ... F_{n_i}^i$ is a sufficient cause for the outcome D and the collection of conjunctions of the form $A_i F_1^i ... F_{n_i}^i$ 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. The central result of this section is essentially that if we have some subset, $X_1, ..., X_m$, of the binary causes $X_1, ..., X_s$ such that there exists an individual for whom intervening to set all of $X_1, ..., X_m$ to 1 ensures that the outcome D is 1 but for whom having all but one of $X_1, ..., X_m$ set to 1 forces the outcome D to be 0 then every sufficient cause representation for D will have some sufficient cause in which all of $X_1, ..., X_m$ are present in the sufficient cause's conjunction. In such cases, since any sufficient cause representation must have the conjunction $X_1 ... X_m$ in its representation, it makes sense to speak of a sufficient cause interaction between $X_1, ..., X_m$.

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). Suppose that $F_1, ..., F_m$ are such that each F_k is either a member of the set of binary causes $\{X_1, ..., X_s\}$ or is the complement of such a member then $F_1, ..., F_m$ 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 $F_1, ..., F_m$ within its conjunction.

Definition 6 requires there to exist within *every* non-redundant minimal sufficient cause representation for D a sufficient cause which has $F_1, ..., F_m$ in its conjunction. Example 2 below makes clear that it does not suffice to merely require that there exists some non-redundant minimal sufficient cause representation for D within which a sufficient cause has $F_1, ..., F_m$ in its conjunction.

EXAMPLE 2. Suppose that the sample space Ω of individuals in the population is given by

$$\Omega_D = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$$

and that

$$D = A_1X_1X_2 \bigvee A_2X_1\overline{X_2} \bigvee A_3X_1X_3 \bigvee A_4X_2\overline{X_3} \bigvee A_5X_1\overline{X_3} \bigvee A_6\overline{X_2}X_3$$

where $A_1 = 1(\omega \in \{\omega_1, \omega_2\}), A_2 = 1(\omega \in \{\omega_1, \omega_3\}), A_3 = 1(\omega \in \{\omega_2, \omega_4\}), A_4 = 1(\omega \in \{\omega_2, \omega_5\}), A_5 = 1(\omega \in \{\omega_3, \omega_6\})$ and $A_6 = 1(\omega \in \{\omega_3, \omega_7\})$. It can be verified that the representation given above is a non-redundant minimal sufficient cause representation for D. Clearly in this non-redundant minimal sufficient cause representation for D, there is a sufficient cause that contains both X_1 and X_2. If however we let $A_7 = 1(\omega = \omega_1)$, then

$$D = A_7X_1 \bigvee A_3X_1X_3 \bigvee A_4X_2\overline{X_3} \bigvee A_5X_1\overline{X_3} \bigvee A_6\overline{X_2}X_3$$

is also a non-redundant minimal sufficient cause representation for D and this representation has no sufficient cause with both X_1 and X_2 in its conjunction.

Corresponding to the definition of a minimal sufficient cause interaction is that of a sufficient cause interaction.

DEFINITION 7 (SUFFICIENT CAUSE INTERACTION). A conjunction of $F_1, ..., F_m$, where each F_k is either a member of the set of binary causes $\{X_1, ..., X_s\}$ or is the complement of such a member, is said exhibit a sufficient cause interaction (or to be irreducible) if within every sufficient cause representation for D there exists some sufficient cause which contains $F_1, ..., F_m$ within its conjunction.

THEOREM 2. The conjunction of $F_1, ..., F_m$ is irreducible if and only if $F_1, ..., F_m$ exhibits a minimal sufficient cause interaction.

We will say that there is a biologic interaction or synergism between the effects of $F_1, ..., F_m$ on D if there exists a sufficient cause for D with $F_1, ..., F_m$ in its conjunction which represents a particular causal mechanism for D. Examples 1 and 2 above suggest 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 a biologic interaction between $F_1, ..., F_m$ is present. In Example 2, it is not possible to distinguish merely from the counterfactual outcomes whether

$$A_1X_1X_2 \bigvee A_2X_1\overline{X_2} \bigvee A_3X_1X_3 \bigvee A_4X_2\overline{X_3} \bigvee A_5X_1\overline{X_3} \bigvee A_6\overline{X_2}X_3$$

or

$$A_7X_1 \bigvee A_3X_1X_3 \bigvee A_4X_2\overline{X_3} \bigvee A_5X_1\overline{X_3} \bigvee A_6\overline{X_2}X_3$$

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 there is a biologic interaction between X_1 and X_2. The presence of biologic interactions 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 biologic interactions will in general require some knowledge of the subject matter in question. 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 biologic interaction and causal mechanisms 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. Although the presence of biologic interactions are sometimes unidentified from the complete set of counterfactual outcomes, they are not always unidentified. If the conjunction of $F_1, ..., F_m$ is irreducible then within every sufficient cause representation for D there exists some sufficient cause which contains $F_1, ..., F_m$ within its conjunction and so there must be some causal mechanism for which $F_1, ..., F_m$ are required; a biologic interaction 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 biologic interactions must be present. Theorem 3 relates this class 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 a biologic interaction must be present.

Before we state these two theorems we will need one additional concept. Some cause I of D is said to be an intermediate variable of other causes of D, $X_1, ..., X_m$ if $I_{x_1...x_m}$, the counterfactual value of I intervening to set $X_1, ..., X_m$ to $x_1, ..., x_m$, is not independent of the values of $x_1, ..., x_m$. We will also need an additional "consistency" assumption: if I is some set of variables such that no variable in I is an intermediate variable between $X_1, ..., X_m$ and D then $D_{X_1=x_1,...,X_m=x_m,I=I(\omega)}(\omega) = D_{X_1=x_1,...,X_m=x_m}(\omega)$.

Theorem 3. Let $X_1, ..., X_m$, be some subset (with the subscripts relabeled if necessary) of $X_1, ..., X_s$ and suppose that none of $X_{m+1}, ..., X_s$ are intermediate variables between $X_1, ..., X_m$ and D. Let Ω denote the sample space for the population and let $D_{x_1...x_m}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_i were set to x_i for $i = 1, ..., m$ then the following two implications hold: (i) if there exists $\omega \in \Omega$ such that $D_{x_1...x_m}(\omega) = 1$ when $x_1 = ... = x_m = 1$ but $D_{x_1...x_m}(\omega) = 0$ for all $x_1, ..., x_m$ such that $\sum_{i=1}^{m} x_i = m - 1$ then $X_1, ..., X_m$ have a sufficient cause interaction; (ii) if $X_1, ..., X_m$ have a sufficient cause interaction then there exists $\omega^* \in \Omega$ and values $x^*_{m+1}, ..., x^*_s$ of $X_{m+1}, ..., X_s$ such that $D_{x_1...x_m x^*_{m+1}...x^*_s}(\omega^*) = 1$ when $x_1 = ... = x_m = 1$ but such that $D_{x_1...x_m x^*_{m+1}...x^*_s}(\omega^*) = 0$ for any $x_1, ..., x_m$ such that $\sum_{i=1}^{m} x_i = m - 1$.

The conditions provided in Theorem 3 have obvious analogues if one or more of $X_1, ..., X_m$ are replaced with their complements. Part (i) of Theorem 3 is of primary interest in this paper; it gives a sufficient condition for a sufficient cause interaction and it will be used below to derive an empirical test for the presence of a sufficient cause interaction.

Remark 1. Theorem 3 and Theorem 4 below require that none of $X_{m+1}, ..., X_s$ be intermediate variables between $X_1, ..., X_m$ and D. This condition is satisfied trivially if $X_1, ..., X_m$ are all of the causes of D under consideration i.e. $X_1, ..., X_s$. The assumption is necessary because it might otherwise be possible that one of the causes $X_{m+1}, ..., X_s$, say X_s, is in fact effectively a conjunction of $X_1, ..., X_m$ or some subset of these variables. In such a case X_s might serve as a proxy for a minimal sufficient cause interaction term and thereby allow for a representation in which the conjunction of $X_1, ..., X_m$ is not present in any sufficient cause. For example, consider three binary causes, X_1, X_2 and X_3, of some binary variable D such that $D_{X_1=x_1,X_2=x_2,X_3=x_3}(\omega) = x_3$ for all $\omega \in \Omega$ and such that $X_{3X_1=x_1,X_2=x_2}(\omega) = x_1x_2$ for all $\omega \in \Omega$ so that X_3 is an intermediate variable between X_1, X_2 and D. In this example we have for any $\omega \in \Omega$ that $D_{X_1=1,X_2=1}(\omega) = 1$ and

$D_{X_1=1,X_2=0}(\omega) = D_{X_1=0,X_2=1}(\omega) = 0$. However, there is no sufficient cause interaction between X_1 and X_2 since $D_{X_1=x_1,X_2=x_2,X_3=x_3}(\omega) = x_3$. Because an intermediate variable between X_1, X_2 and D is being considered in the sufficient cause representations for D no sufficient cause interaction is manifest. If X_3 were not being considered in the sufficient cause representations we would have that $D_{X_1=x_1,X_2=x_2}(\omega) = x_1x_2$ and the counterfactual conditions $D_{X_1=1,X_2=1}(\omega) = 1$ and $D_{X_1=1,X_2=0}(\omega) = D_{X_1=0,X_2=1}(\omega) = 0$ would imply the presence of a sufficient cause interaction between X_1 and X_2. It is evident from this example that intermediate variables can obscure the presence of a sufficient cause interaction. When the causes $X_1, ..., X_s$ are such that none of these variables is a cause of another, these difficulties do not arise.

Remark 2. Note that under Definition 7, the presence of a sufficient cause interaction may depend upon the context of which other causes $X_{m+1}, ..., X_s$ are being considered in the sufficient cause representations. However, the condition that there exists $\omega \in \Omega$ such that $D_{x_1...x_m}(\omega) = 1$ when $x_1 = ... = x_m = 1$ but $D_{x_1...x_m}(\omega) = 0$ for all $x_1, ..., x_m$ such that $\sum_{i=1}^{m} x_i = m - 1$ does not make reference to $X_{m+1}, ..., X_s$ and thus provides a condition for a sufficient cause interaction that is not dependent on the context. If this condition holds then the conjunction $X_1...X_m$ will be present in any sufficient cause representation regardless of which other causes $X_{m+1}, ..., X_s$ are being considered in the sufficient cause representations so long as the these other causes of D, $X_{m+1}, ..., X_s$, are not themselves effects of $X_1, ..., X_m$.

Theorem 3 suggests a very natural empirical condition for detecting the presence of a sufficient cause interaction. Some discussion with regard to constructing statistical tests related to this condition is given following the statement of Theorem 3. In Section 5, the condition stated in Theorem 3 is related explicitly to statistical tests arising from generalized linear models.

Theorem 4. Let $X_1, ..., X_m$, be some subset (with the subscripts relabeled if necessary) of $X_1, ..., X_s$ and suppose that none of $X_{m+1}, ..., X_s$ are intermediate variables between $X_1, ..., X_m$ and D. Let Ω denote the sample space for the population and let $D_{x_1...x_m}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_i were set to x_i for $i = 1, ..., m$. Let C be any set of variables which suffices to control for the confounding of the causal effects of $X_1, ..., X_m$ on D i.e. such that $D_{x_1...x_m} \coprod \{X_1, ..., X_m\}|C$ then if for any value c of C we have that

$$\begin{aligned} &E(D|X_1 = 1, ..., X_m = 1, C = c) \\ &\quad - E(D|X_1 = 0, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\ &\quad - E(D|X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\ &\quad - ... - \\ &\quad - E(D|X_1 = 1, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 0, C = c) > 0 \end{aligned}$$

then $X_1, ..., X_m$ have a sufficient cause interaction.

As with Theorem 3, the condition provided in Theorem 4 has obvious analogues if one or more of $X_1, ..., X_m$ are replaced with their complements. If the set of confounding variables C consists of a small number of binary or categorical variables then it may be possible to use the 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 $pr(D = 1|X_1, ..., X_m, C)$ using a binomial or Poisson regression model with a linear link (Greenland, 1991; Wacholder, 1986; Zou, 2004; Greenland, 2004; Spiegelman and Hertzmark, 2005). Such approaches are discussed in more detail in Section 5. For case-control studies it will be necessary to use an adapted set of modeling techniques (Wild, 1991; Wacholder, 1996; Greenland, 2004).

Remark 3. Note that even if the condition of Theorem 4 is met it will in general be difficult to identify exactly which sufficient cause has $X_1, ..., X_m$ in its conjunction. If the causes of D in the set $\{X_1, ..., X_s\}$ other than $\{X_1, ..., X_m\}$ suffices to control for the confounding of the causal effect of $X_1, ..., X_m$ on D then some progress can be made. From the proof of Theorems 3 and 4 it follows that if C can be chosen to be this set then at least one of the sufficient causes with $X_1, ..., X_m$ in its conjunction will also have in its conjunction some subset of the restrictions imposed on this confounding set by $C = c$. But it is important to note that in any particular non-redundant minimal sufficient cause representation for D there may be multiple sufficient causes with $X_1, ..., X_m$ in their conjunction; also different non-redundant minimal sufficient cause representations for D may have $X_1, ..., X_m$ in the conjunctions of different sufficient causes. Finally, even if it is the case that the conditions of Theorem 4 fail there might still be a sufficient cause with $X_1, ..., X_m$ in its conjunction along with some subset of the restrictions imposed on the confounding set by $C = c$. The condition provided in Theorem 4 is sufficient but not necessary for the presence of a sufficient cause interaction.

4. Sufficient Cause Interactions and Monotonic Effects

Sometimes it may be known that a certain cause has an effect on an outcome that is always in a particular direction, always positive or always negative. In the context of testing for sufficient cause interactions, knowledge of the "monotonicity" of certain causes will allow for the construction of more powerful statistical tests than those constructed from Theorem 4. We begin with the definition of a monotonic effect.

Definition 8 (Monotonic Effect). We will say that $X_1, ..., X_m$ have positive monotonic effects on D if for all individuals ω we have $D_{x_1...x_m}(\omega) \geq D_{x'_1...x'_m}(\omega)$ whenever $x_i \geq x'_i$ for $i = 1, ..., m$. Similarly, we will say that say that $X_1, ..., X_m$ have negative monotonic effects on D if for all individuals ω we have $D_{x_1...x_m}(\omega) \leq D_{x'_1...x'_m}(\omega)$ whenever $x_i \geq x'_i$ for $i = 1, ..., m$.

The definition of a monotonic effect essentially requires that some intervention or set of interventions either increase or decrease some outcome D not merely on average over the population but rather for every individual in that population regardless of the other interventions taken. 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. A more general definition of monotonic effects can be given as follows: the variables $X_1, ..., X_m$ are said to have positive monotonic effects on D relative to $X_1, ..., X_s$ if for all individuals ω and all values of $x_{m+1}, ..., x_s$ we have $D_{x_1...x_m x_{m+1}...x_s}(\omega) \geq D_{x'_1...x'_m x_{m+1}...x_s}(\omega)$ whenever $x_i \geq x'_i$ for $i = 1, ..., m$; however, the less general definition given in Definition 8 suffices for the purposes of this paper. 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 within the directed acyclic graph causal framework.

The main result in this section is an analogue, under the additional assumption of monotonic effects, to Theorem 4. Theorem 6 provides this analogue. Because of the assumption of monotonic effects, Theorem 6 gives a weaker condition to be tested than did Theorem 4 and thus yields more powerful statistical tests. Theorem 6 is preceded by Theorem 5 which develops the counterfactual condition upon which Theorem 6 is based. However, before stating Theorems 5 and 6 we introduce one further concept, that of a subordinate set. Theorems 5 and 6 allow for more powerful statistical tests essentially because, under the assumption of monotonic effects, it is possible to add to the conditions of Theorems 3 and 4 several terms corresponding to counterfactual outcomes for combinations of causes fixed by a subordinate set.

Definition 9 (Subordinate Set). Let $\mathcal{U}(m) = \{(x_1, ..., x_m) \in \{0,1\}^m : \sum_{i=1}^{m} x_i = m - 1\}$ and let $\mathcal{Q}(m) = \{(x_1, ..., x_m) \in \{0,1\}^m : \sum_{i=1}^{m} x_i = m - 2\}$ then we say that $\mathcal{S}$ is an subordinate set of order m if $\mathcal{S}$ consists of $m-1$ members of $\mathcal{Q}$ such that for any $m-1$ distinct members of $\mathcal{U}$, $u_1, ..., u_{m-1}$, the members of $\mathcal{S}$ can be ordered $s_1, ..., s_{m-1}$ so that $s_i \leq u_i$ for $i = 1, ..., m-1$.

The set $\mathcal{U}(m)$ has m members and the set $\mathcal{Q}(m)$ has has $\binom{m}{2}$ members. The requirement that each member $(x_1, ..., x_m)$ of the set $\mathcal{Q}(m)$ be such that $\sum_{i=1}^{m} x_i = m-2$ is simply that each member of $\mathcal{Q}(m)$ have $m-2$ x_i's with the value 1 and two x_i's with the value of 0 and the requirement that each member $(x_1, ..., x_m)$ of the set $\mathcal{U}(m)$ be such that $\sum_{i=1}^{m} x_i = m-1$ is simply that for some j, $x_j = 0$ and for $i \neq j$ we have that $x_i = 1$. There will in general many possible subordinate sets $\mathcal{S}$ of a particular order. Although there is only one subordinate set of order 2:

$$\{(0,0)\}$$

it can be shown that there are three subordinate sets of order 3:

$$\{(1,0,0),(0,1,0)\}$$
$$\{(1,0,0),(0,0,1)\}$$
$$\{(0,1,0),(0,0,1)\}$$

and that there are sixteen subordinate sets of order 4:

$$\{(1,1,0,0),(1,0,1,0),(0,1,1,0)\}$$
$$\{(1,1,0,0),(1,0,1,0),(0,1,0,1)\}$$
$$\{(1,1,0,0),(1,0,1,0),(0,0,1,1)\}$$
$$\{(1,1,0,0),(1,0,0,1),(0,1,1,0)\}$$
$$\{(1,1,0,0),(1,0,0,1),(0,1,0,1)\}$$
$$\{(1,1,0,0),(1,0,0,1),(0,0,1,1)\}$$
$$\{(1,1,0,0),(0,1,1,0),(0,0,1,1)\}$$
$$\{(1,1,0,0),(0,1,0,1),(0,0,1,1)\}$$
$$\{(1,0,1,0),(1,0,0,1),(0,1,1,0)\}$$
$$\{(1,0,1,0),(1,0,0,1),(0,1,0,1)\}$$
$$\{(1,0,1,0),(1,0,0,1),(0,0,1,1)\}$$
$$\{(1,0,1,0),(0,1,1,0),(0,1,0,1)\}$$
$$\{(1,0,1,0),(0,1,0,1),(0,0,1,1)\}$$
$$\{(1,0,0,1),(0,1,1,0),(0,1,0,1)\}$$
$$\{(1,0,0,1),(0,1,1,0),(0,0,1,1)\}$$
$$\{(0,1,1,0),(0,1,0,1),(0,0,1,1)\}$$

For $m = 4$, an example of a set that it is not subordinate is $\mathcal{S}^* = \{(1,1,0,0),(1,0,1,0),(1,0,0,1)\}$ because three distinct members of $\mathcal{U}(4) = \{(x_1, x_2, x_3, x_4) : \sum_{i=1}^{4} x_i = 3\}$ can be chosen as $\{(0,1,1,1),(1,1,1,0),(1,1,0,1)\}$ but for no member s of $\mathcal{S}^*$ is it true that $s \leq (0,1,1,1)$ since every member of $\mathcal{S}^*$ has value 1 in the first dimension of the vector (x_1, x_2, x_3, x_4). To keep the exposition relatively simple, we will for the remainder of the paper restrict our examples to those requiring only subordinate sets of order not more than 3. The definition of a subordinate set allows us to state Theorem 5 which provides an analogue, under the assumption of monotonic effects, to Theorem 3.

Theorem 5. Let $X_1, ..., X_m$ be some subset (with the subscripts relabeled if necessary) of $X_1, ..., X_s$ and suppose that $X_1, ..., X_m$ have monotonic effects on D and that none of $X_{m+1}, ..., X_s$

are intermediate variables between $X_1, ..., X_m$ and D. Let Ω denote the sample space for the population, let $D_{x_1...x_m}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_i were set to x_i for $i = 1, ..., m$ and let $\mathcal{U} = \{(x_1, ..., x_m) \in \{0,1\}^m : \sum_{i=1}^m x_i = m-1\}$. If there exists an ω such that $D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) > 0$ for some subordinate set $\mathcal{S}$ of order m then $X_1, ..., X_m$ have a sufficient cause interaction.

REMARK 4. In the proof of Theorem 5 it was shown that under the assumption of monotonic effects, if for some choice of a subordinate set $\mathcal{S}$ we have that

$$D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) > 0 \quad (1)$$

then we must also have that

$$D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) > 0. \quad (2)$$

Clearly the converse is also true i.e. condition (2) implies condition (1) since the expression $\sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega)$ is non-negative. Condition (2) is exactly equivalent to the condition in part (i) of Theorem 3 that $D_{x_1...x_m}(\omega) = 1$ when $x_1 = ... = x_m = 1$ but $D_{x_1...x_m}(\omega) = 0$ for any $x_1...x_m$ such that $\sum_{i=1}^m x_i = m-1$. The proof of Theorem 5 thus demonstrates that, with the assumption of monotonic effects, condition (1) is exactly equivalent to the condition in part (i) of Theorem 3. We will see below however that using condition (1) instead of condition (2) allows for the construction of more powerful statistical tests for the presence of a sufficient cause interaction.

In the case of a two-way sufficient cause interaction, the counterfactual condition of Theorem 5 is simply:

$$D_{11}(\omega) - D_{10}(\omega) - D_{01}(\omega) + D_{00}(\omega) > 0.$$

For the case of $m = 3$, there are three choices for the subordinate set $\mathcal{S}$ and thus there exists a three-way sufficient cause interaction if any of the following three conditions hold:

$$\begin{aligned}
D_{111}(\omega) - D_{110}(\omega) - D_{101}(\omega) - D_{011}(\omega) + D_{100}(\omega) + D_{010}(\omega) &> 0 \\
D_{111}(\omega) - D_{110}(\omega) - D_{101}(\omega) - D_{011}(\omega) + D_{100}(\omega) + D_{001}(\omega) &> 0 \\
D_{111}(\omega) - D_{110}(\omega) - D_{101}(\omega) - D_{011}(\omega) + D_{010}(\omega) + D_{001}(\omega) &> 0
\end{aligned}$$

Theorem 6 is an empirical test of the counterfactual conditions provided by Theorem 5. The form of the proof of Theorem 6 is essentially equivalent to that of Theorem 4 and is therefore suppressed.

THEOREM 6. Let $X_1, ..., X_m$, be some subset (with the subscripts relabeled if necessary) of $X_1, ..., X_s$ and suppose that $X_1, ..., X_m$ have monotonic effects on D and that none of $X_{m+1}, ..., X_s$ are intermediate variables between $X_1, ..., X_m$ and D. Let Ω denote the sample space for the population and let $D_{x_1...x_m}(\omega)$ be the counterfactual value of D for $\omega \in \Omega$ if X_i were set to x_i for $i = 1, ..., m$. Let C be any set of variables which suffice to control for the confounding of the causal effects of $X_1, ..., X_m$ on D i.e. such that $D_{x_1...x_m} \coprod \{X_1, ..., X_m\} | C$ and let $\mathcal{S}$ be any subordinate set of order m then if for any value c of C we have that

$$\begin{aligned}
&E(D|X_1 = 1, ..., X_m = 1, C = c) \\
&\quad - E(D|X_1 = 0, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\
&\quad - E(D|X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\
&\quad - ... - \\
&\quad - E(D|X_1 = 1, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 0, C = c) \\
&\quad + \sum_{(x_1,...,x_m)\in\mathcal{S}} E(D|X_1 = x_1, ..., X_m = x_m, C = c) > 0
\end{aligned}$$

then $X_1, ..., X_m$ have a sufficient cause interaction.

The result given in Theorem 6 in the special case of $m = 2$ with no confounding factors is stated explicitly and proved by Rothman and Greenland (1998); this special case is also anticipated elsewhere (Koopman, 1981; Darroch and Borkent, 1994). Like Theorems 3 and 4, Theorems 5 and 6 have analogues if one or more of $X_1, ..., X_m$ are replaced with their complements and if one or more of $X_1, ..., X_m$ have negative rather than positive monotonic effects on D. The condition of Theorem 6 is clearly weaker than that of Theorem 4 which did not assume monotonic effects. This is evident because of the addition of the term $\sum_{(x_1,...,x_m)\in\mathcal{S}} E(D|X_1 = x_1, ..., X_m = x_m, C = c) \geq 0$. As with the condition of Theorem 4, the condition in Theorem 6 can be tested with t-test like test statistics or using various statistical models. It is these statistical models and the relation between sufficient cause interactions and interaction terms in standard statistical models that is the topic of the following section.

5. Sufficient Cause Interactions and Interaction Terms in Statistical Models

Theorems 4 and 6 provide conditions which can be empirically tested to draw inferences about the presence of sufficient cause interactions. If a sufficient cause interaction between $X_1, ..., X_m$ is present, then in any sufficient cause representation for the outcome there must exist a sufficient cause in which $X_1, ..., X_m$ are all present - the causal mechanisms for the outcome must be such that $X_1, ..., X_m$ all participate in the same causal mechanism. 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. However, 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. Other modeling approaches must be used. The conditions provided in Theorems 4 and 6 are given in terms of differences between various probabilities. This suggests a Bernoulli regression model with linear link as the natural choice to test the conditions of these two theorems. Unfortunately, when continuous covariates are included in Bernoulli regressions with linear link, the convergence properties of maximum likelihood estimates are generally poor (Wacholder, 1986). If, however, all covariates in a Bernoulli regression with linear link are binary or categorical, it is possible to use a saturated model for the outcome and maximum likelihood regression estimates will always converge. We intend to address in future work more sophisticated modeling issues and approaches with regard to testing for sufficient cause interactions. Here however we will simply compare, within a regression framework, the tests arising from Theorems 4 and 6 with standard tests for statistical interactions. For simplicity we will assume that the causal effects of $X_1, ..., X_m$ are unconfounded. The substance of the remarks below are not altered in the case of one or more binary or categorical confounding variables.

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

$$pr(D = 1|X_1 = x_1, X_2 = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

We will use $p_{x_1 x_2}$ as a shorthand for $pr(D = 1|X_1 = x_1, X_2 = x_2)$. 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, then Theorem 6 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 $\beta_3 > 0$ 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.

Let us 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:

$$pr(D = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3) = \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_1 x_2+\beta_5 x_1 x_3+\beta_6 x_2 x_3+\beta_7 x_1 x_2 x_2.$$

Once again we will use the shorthand $p_{x_1 x_2 x_3} = pr(D = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3)$. 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, Theorem 6 states that X_1, X_2 and X_3 exhibit a sufficient cause interaction if any of the 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 apply Theorem 4 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. 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 will perhaps 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.

Several 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 method concerns the cases in which a biologic interaction 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 a biologic interaction to be present. A biologic interaction can be present even if the conditions of Theorems 3-6 do not hold. Such biologic interactions cannot be identified from data.

In future work we intend to find application for the empirical tests derived in this paper and we intend also to pursue various modeling approaches to testing the conditions given in Theorems 4 and 6 in the presence of continuous confounding variables.

Appendix

Proof of Theorem 1.
For $G_i = F_1^i ... F_{n_i}^i$ we construct the corresponding binary variable A_i as follows. Recall that each condition of the form $F_k^i = 1$ places a restriction on one of $X_1, ..., X_s$, that it be either 1 or 0. Let $A_i(\omega) = 1$ if $D_{x_1 ... x_s}(\omega) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1^i ... F_{n_i}^i = 1$$

and if there does not exist a j such that $D_{x_1 ... x_s}(\omega) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1^i ... F_{j-1}^i F_{j+1}^i ... F_{n_i}^i = 1.$$

Otherwise, let $A_i(\omega) = 0$. We will show that

$$D_{x_1 ... x_s}(\omega) = \bigvee_i A_i(\omega) g_i(x_1, ..., x_s).$$

Consider ω and $x_1, ..., x_s$ such that

$$\bigvee_i A_i(\omega) g_i(x_1, ..., x_s) = 1.$$

Then there exists an i such that $A_i(\omega) g_i(x_1, ..., x_s) = 1$. Since $A_i(\omega) = 1$ we have that $D_{x_1 ... x_s}(\omega) = 1$ whenever $x_1, ..., x_s$ are such that $F_1^i ... F_{n_i}^i = 1$; and since $g_i(x_1, ..., x_s) = 1$ we have $(X_1, ..., X_s) = (x_1, ..., x_s)$ implies that $F_1^i ... F_{n_i}^i = 1$ and thus we have that $D_{x_1 ... x_s}(\omega) = 1$. Now we must show that if $D_{x_1 ... x_s}(\omega) = 1$ then there exists an i such that $A_i(\omega) g_i(x_1, ..., x_s) = 1$. The potential outcome $D_{x_1 ... x_s}(\omega)$ is a function of $(\omega, x_1, ..., x_s)$. Let $(\omega^*, x_1^*, ..., x_s^*)$ be such that $D_{x_1^* ... x_s^*}(\omega^*) = 1$. Consider the ordered set $(x_1^*, ..., x_s^*)$. If for any j,

$$x_1 = x_1^*, ..., x_{j-1} = x_{j-1}^*, x_{j+1} = x_{j+1}^*, ..., x_m = x_m^* \Rightarrow D_{x_1 ... x_s}(\omega^*) = 1$$

remove x_j^* from $(x_1^*, ..., x_s^*)$. Continue to remove those x_j^* from this set which are not needed to maintain the implication $D_{x_1 ... x_s}(\omega^*) = 1$. Suppose the set that remains is $(x_{h_1}^*, ..., x_{h_u}^*)$. We then have that

$$x_{h_1} = x_{h_1}^*, ..., x_{h_u} = x_{h_u}^* \Rightarrow D_{x_1 ... x_s}(\omega^*) = 1 \quad (3)$$

and that for no j,

$$x_{h_1} = x^*_{h_1}, ..., x_{h_{j-1}} = x^*_{h_{j-1}}, x_{h_{j+1}} = x^*_{h_{j+1}}, ..., x_{h_u} = x^*_{h_u} \Rightarrow D_{x_1...x_s}(\omega^*) = 1. \quad (4)$$

Define F_j as the indicator $F_j = 1_{(X_{h_j} = x^*_{h_j})}$, then for some i, $G_i = F_1...F_u$. Since (3) and (4) hold, we must have that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1...F_u = 1$$

and that for no j is it the case that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1...F_{j-1}F_{j+1}...F_u = 1.$$

Thus $A_i(\omega^*) = 1$. Since $(X_1, ..., X_s) = (x^*_1, ..., x^*_s) \Rightarrow X_{h_1} = x^*_{h_1}, ..., X_{h_u} = x^*_{h_u} \Rightarrow F_1...F_u = 1$ we have that $g_i(x^*_1, ..., x^*_s) = 1$ and so $A_i(\omega^*)g_i(x^*_1, ..., x^*_s) = 1$ and so $\bigvee_i A_i(\omega^*)g_i(x^*_1, ..., x^*_s) = 1$. We have thus shown $D_{x_1...x_s}(\omega) = \bigvee_i A_i(\omega)g_i(x_1, ..., x_s)$. From this it also immediately follows that $D(\omega) = \bigvee_i A_i(\omega)F^i_1(\omega)...F^i_{n_i}(\omega)$ since

$$D(\omega) = D_{X_1(\omega)...X_s(\omega)}(\omega) = \bigvee_i A_i(\omega)g_i\{X_1(\omega), ..., X_s(\omega)\} = \bigvee_i A_i(\omega)F^i_1(\omega)...F^i_{n_i}(\omega).$$

Proof of Theorem 2.
If $F_1...F_n$ is irreducible then within any sufficient cause representation $\bigvee_i A_iF^i_1...F^i_{n_i}$ there exists some sufficient cause which contains within its conjunction $F_1, ..., F_n$ 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 $F_1, ..., F_n$ in its conjunction. If $F_1...F_n$ is not irreducible then there exists some representation $\bigvee_i A_iF^i_1...F^i_{n_i}$ such that no sufficient cause $A_iF^i_1...F^i_{n_i}$ contains within its conjunction $F_1, ..., F_n$. This representation $\bigvee_i A_iF^i_1...F^i_{n_i}$ can be made into a non-redundant minimal sufficient cause representation by iteratively discarding the components of each conjunction $A_iF^i_1...F^i_{n_i}$ which are not necessary to preserve the implication $A_iF^i_1...F^i_{n_i} \Rightarrow D = 1$ and then iteratively discarding any redundant minimal sufficient causes. Clearly no sufficient cause of this resulting non-redundant minimal sufficient causation representation will contain $F_1, ..., F_n$ within its conjunction.

Proof of Theorem 3.
Suppose that $X_1, ..., X_m$ do not have a minimal sufficient cause interaction then there exists a non-redundant minimal sufficient cause representation $\bigvee_i A_iF^i_1...F^i_{n_i}$ such that there is no sufficient cause within the representation which contains $X_1, ..., X_m$ in its conjunction. Note that $X_{m+1}, ..., X_s$ are the members of $\{X_1, ..., X_s\}$ other than $X_1, ..., X_m$. Consider any $\omega \in \Omega$ such that $D_{x_1...x_m}(\omega) = 0$ whenever $\sum_{i=1}^m x_i = m - 1$. Suppose $X_{m+1}(\omega) = x_{m+1}, ..., X_s(\omega) = x_s$. Define $J_{m+1}, ..., J_s$ by $J_{m+1} = 1(X_{m+1} = x_{m+1}), ..., J_s = 1(X_s = x_s)$. For every $G_i = F^i_1...F^i_{n_i}$ for which the F^i_k's consist only of some subset of the elements of $X_1, ..., X_{j-1}, X_{j+1}, ...X_m, J_{m+1}, ..., J_s$ we must have $A_i(\omega) = 0$ since for each j, $D_{x_1...x_m}(\omega) = 0$ when $x_j = 0$ and $x_i = 1$ for $i \neq j$. By assumption there was no sufficient cause G_i within the representation which included $X_1, ..., X_m$ in its conjunction. Thus for every $G_i = F^i_1...F^i_{n_i}$ for which the F^i_k's consist only of some subset of the elements of $X_1, ..., X_m, J_{m+1}, ..., J_s$ we must have $A_i(\omega) = 0$ and so we have that $D_{X_1=1,...,X_m=1,X_{m+1}=x_{m+1},...,X_s=x_s}(\omega) = 0$. Furthermore, since none of $X_{m+1}, ..., X_s$ are intermediate variables between $X_1, ..., X_m$ and D we have that

$$D_{X_1=1,...,X_m=1}(\omega) = D_{X_1=1,...,X_m=1,X_{m+1}=x_{m+1},...,X_s=x_s}(\omega) = 0.$$

There thus exists no $\omega \in \Omega$ such that $D_{X_1=1,...,X_m=1}(\omega) = 1$ but $D_{x_1...x_m}(\omega) = 0$ whenever $\sum_{i=1}^m x_i = m-1$. We now prove the second proposition of the Theorem. Suppose that $X_1, ..., X_m$ do have a minimal sufficient cause interaction. The representation for D given in the proof of Theorem 1 may be reduced to a non-redundant minimal sufficient cause representation by iteratively excluding any unnecessary components from sufficient causes that are not minimally sufficient and then eliminating those minimal sufficient causes which are redundant. Let $\bigvee_i A_i F_1^i...F_{n_i}^i$ be the resulting non-redundant minimal sufficient cause representation. Since $X_1, ..., X_m$ have a minimal sufficient cause interaction there exists within this non-redundant minimal sufficient cause representation a sufficient cause which has $X_1, ..., X_m$ in its conjunction. Suppose $G_l = X_1...X_m H_1...H_u$ where each of $H_1, ..., H_u$ are members of the set $\{X_{m+1}, ..., X_s\}$ or complements of such members. Let K be the set of elements in $\{X_{m+1}, ..., X_s\}$ which are not or whose complements are not in the conjunction $G_l = X_1...X_m H_1...H_u$. Since the minimal sufficient cause representation $\bigvee_i A_i F_1^i...F_{n_i}^i$ is non-redundant there exists $\omega^* \in \Omega$ and some value K^* of K such that $\omega = \omega^*$ fixes $A_l(\omega^*) = 1$ and such that $A_l X_1...X_m H_1...H_u$ is the only sufficient cause to take the value 1 in the representation $\bigvee_i A_i F_1^i...F_{n_i}^i$ when $\omega = \omega^*$ and $\{X_1, ..., X_s\}$ take the values corresponding to $X_1, ..., X_m, H_1, ..., H_u$ and K^*. Let $x_{m+1}^*, ..., x_s^*$ be the values of $x_{m+1}, ..., x_s$ which correspond to $H_1, ..., H_u$ and K^*. We then have that $D_{1...1x_{m+1}^*...x_s^*}(\omega^*) = 1$. We will show that $D_{x_1...x_m x_{m+1}^*...x_s^*}(\omega^*) = 0$ whenever $\sum_{i=1}^m x_i = m-1$. Suppose that for some j, $D_{x_1...x_m x_{m+1}^*...x_s^*}(\omega^*) = 1$ when $x_j = 0$ and $x_i = 1$ for $i = 1, ..., j-1, j+1, ...m$ then by the consistency assumption we would have that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$X_1...X_{j-1}\overline{X_j}X_{j+1}...X_m H_1...H_u = 1, K = K^*$$

and that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$X_1...X_{j-1}X_j X_{j+1}...X_m H_1...H_u = 1, K = K^*.$$

From this it follows that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$X_1...X_{j-1}X_{j+1}...X_m H_1...H_u = 1, K = K^*$$

But since $A_l X_1...X_m H_1...H_u$ is the only minimal sufficient cause to take the value 1 in the non-redundant minimal sufficient causation representation $\bigvee_i A_i F_1^i...F_{n_i}^i$ when $\omega = \omega^*, X_1...X_m H_1...H_u = 1, K = K^*$ it would then also follow that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$X_1...X_{j-1}X_{j+1}...X_m H_1...H_u = 1$$

but this contradicts $A_l(\omega^*) = 1$ in the construction of the variables A_i's given in Theorem 1 since for the A_i corresponding to $G_i = F_1^i...F_{n_i}^i$ we have that $A_i(\omega^*) = 1$ if and only if $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1^i...F_{n_i}^i = 1$$

and if there does not exist a j such that $D_{x_1...x_s}(\omega^*) = 1$ whenever $x_1, ..., x_s$ are such that

$$F_1^i...F_{j-1}^i F_{j+1}^i...F_{n_i}^i = 1.$$

From this it follows that we must have that $D_{x_1...x_m x_{m+1}^*...x_s^*}(\omega^*) = 0$ whenever $\sum_{i=1}^m x_i = m-1$. There thus exists an ω^* such that $D_{1...1x_{m+1}^*...x_s^*}(\omega^*) = 1$ but such that $D_{x_1...x_m x_{m+1}^*...x_s^*}(\omega^*) = 0$ whenever $\sum_{i=1}^m x_i = m-1$.

Proof of Theorem 4.

We prove the contrapositive. Suppose there were no sufficient cause interaction between $X_1, ..., X_m$ then by Theorem 3 it would follow that there is no $\omega \in \Omega$ such that $D_{1...1}(\omega) = 1$ but such that $D_{x_1...x_m}(\omega) = 0$ whenever $\sum_i x_i = m - 1$. From this it follows that for all $\omega \in \Omega$ we have $D_{1...1}(\omega) - D_{01...1}(\omega) - ... - D_{1...10}(\omega) \leq 0$ and so $E\{D_{1...1}(\omega) - D_{01...1}(\omega) - ... - D_{1...10}(\omega)|C\} \leq 0$. Since $D_{x_1...x_m} \coprod \{X_1, ..., X_m\}|C$ we have that

$$
\begin{aligned}
&E(D|X_1 = 1, ..., X_m = 1, C = c) \\
&\qquad - E(D|X_1 = 0, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\
&\qquad - E(D|X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c) \\
&\qquad - ... - \\
&\qquad - E(D|X_1 = 1, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 0, C = c) \\
&= E\{D_{1...1}(\omega)|X_1 = 1, ..., X_m = 1, C = c\} \\
&\qquad - E\{D_{011...1}(\omega)|X_1 = 0, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c\} \\
&\qquad - E\{D_{101...1}(\omega)|X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{m-1} = 1, X_m = 1, C = c\} \\
&\qquad - ... - \\
&\qquad - E\{D_{1...10}(\omega)|X_1 = 1, X_2 = 1, X_3 = 1, ..., X_{m-1} = 1, X_m = 0, C = c\} \\
&= E\{D_{1...1}(\omega) - D_{01...1}(\omega) - ... - D_{1...10}(\omega)|C\} \leq 0.
\end{aligned}
$$

This completes the proof.

Proof of Theorem 5.
Suppose that for some $\omega \in \Omega$ we have that

$$D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) \leq 0.$$

If $D_{1...1}(\omega) = 0$ then $D_{x_1...x_m}(\omega) = 0$ for all $x_1, ..., x_m$ since $X_1, ..., X_m$ have monotonic effects on D and so

$$D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) = 0.$$

If $D_{1...1}(\omega) \neq 0$ then we must have that $D_{1...1}(\omega) = 1$ and since $D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) \leq 0$ there must exist some $(x'_1, ..., x'_m) \in \mathcal{U}$ such that $D_{x'_1...x'_m}(\omega) = 1$. Let $\mathcal{U}' = \mathcal{U} \backslash (x'_1, ..., x'_m)$. For any choice of the subordinate set $\mathcal{S}$ we have that

$$
\begin{aligned}
&D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) \\
&\qquad = D_{1...1}(\omega) - D_{x'_1...x'_m}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}'} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega).
\end{aligned}
$$

Now $D_{1...1}(\omega) - D_{x'_1...x'_m}(\omega) = 1 - 1 = 0$ and furthermore

$$-\sum_{(x_1,...,x_m)\in\mathcal{U}'} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) \leq 0$$

since each of the two sums has $m - 1$ terms and since, because $\mathcal{S}$ is a subordinate set, each term in the sum over $\mathcal{U}'$ can be matched with a term in the sum over $\mathcal{S}$ so that, because of the assumption that $X_1, ..., X_m$ have monotonic effects on D, the term in the sum over $\mathcal{U}'$ will be at least as large as the term in the sum over $\mathcal{S}$. Thus we have that

$$D_{1...1}(\omega) - \sum_{(x_1,...,x_m)\in\mathcal{U}} D_{x_1...x_m}(\omega) + \sum_{(x_1,...,x_m)\in\mathcal{S}} D_{x_1...x_m}(\omega) \leq 0.$$

We have shown that if

$$D_{1\ldots1}(\omega) - \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{U}} D_{x_1\ldots x_m}(\omega) \leq 0$$

then

$$D_{1\ldots1}(\omega) - \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{U}} D_{x_1\ldots x_m}(\omega) + \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{S}} D_{x_1\ldots x_m}(\omega) \leq 0$$

for any choice of a subordinate set $\mathcal{S}$. From this it follows that if for some choice of a subordinate set $\mathcal{S}$ we have that

$$D_{1\ldots1}(\omega) - \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{U}} D_{x_1\ldots x_m}(\omega) + \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{S}} D_{x_1\ldots x_m}(\omega) > 0$$

then we must also have that

$$D_{1\ldots1}(\omega) - \sum\nolimits_{(x_1,\ldots,x_m)\in\mathcal{U}} D_{x_1\ldots x_m}(\omega) > 0$$

and so by Theorem 3, $X_1, ..., X_m$ have a sufficient cause interaction.

References

Darroch, J.N. and Borkent, M. (1994). Synergism, attributable risk and interaction for two binary exposure factors. *Biometrika* **81**, 259-270.

Davidson, D. (1980). Essays on actions and events. Oxford University Press, Oxford.

Flanders, D. (2006). Sufficient-component cause and potential outcome models. *Eur. J. Epidemiol*, in press.

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.

Hume, D. (1748). *An Enquiry Concerning Human Understanding.* Reprinted Open Court Press, LaSalle, IL, 1958.

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

Lewis, D. (1973a). Causation. *Journal of Philosophy.* **70** 556-567.

Lewis, D. (1973b). *Counterfactuals.* Harvard University Press, Cambridge.

Lewis, D. (1986). Events. In Philosophical Paper, Vol. 2, D. Lewis. Oxford University Press, New York.

Mantel, N., Brown, C. and Byar, D.P. (1977). Tests for homogeneity of effect in an epidemiologic investigation. *Am. J. Epidemiol.* **106**, 125-129.

Neyman, J. (1923). Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, Trans.) in *Statistical Science* **5**, 463–472.

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. *Mathematical Modelling* **7** 1393-1512.

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. *Computers and Mathematics with Applications* **14** 923-945.

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

Rothman, K.J. and Greenland, S. (1998). Modern Epidemiology. Philadelphia, PA: Littincott-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. 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. *Submitted.*

VanderWeele, T.J. and Robins, J.M. (2006b). Signed directed acyclic graphs for causal inference. *Submitted.*

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.