Statistica Sinica **31** (2021), 2195-2212
doi:https://doi.org/10.5705/ss.202019.0207

SUFFICIENT CAUSE INTERACTIONS FOR CATEGORICAL AND ORDINAL OUTCOMES

Jaffer M. Zaidi and Tyler J. VanderWeele

Harvard University

Abstract: The sufficient cause model is extended from binary to categorical and ordinal outcomes to formalize the concept of sufficient cause interaction and synergism in this setting. This extension allows us to derive counterfactual and empirical conditions for detecting the presence of sufficient cause interactions for ordinal and categorical outcomes. Some of these conditions are entirely novel in that they cannot be derived from the sufficient cause model for binary outcomes. These empirical conditions enable researchers to determine whether two exposures display synergism for an ordinal or a categorical outcome. Likelihood ratio tests that use these derived empirical conditions are developed to infer sufficient cause interaction for ordinal and categorical outcomes. Lastly, we apply these likelihood ratio tests to detect sufficient cause interaction between two major resistance mutations in the development of HIV drug resistance to Etravirine.

Key words and phrases: Interaction, ordinal outcome, sufficient cause.

1. Introduction

In this study, we extend the sufficient cause model, defined originally for binary outcomes, to categorical and ordinal outcomes, and derive the associated empirical and counterfactual conditions associated with sufficient cause interaction. The sufficient cause framework represents causation as a collection of causal mechanisms, called sufficient causes. A single sufficient cause is constituted of one or more component causes, such that when all components of the sufficient cause are present, they together inevitably bring about the outcome. The first crude sufficient cause model appeared in (Cayley (1853)). Rothman popularized the sufficient cause model in epidemiology, presenting a graphical schematic that often appears in introductory epidemiology texts (Rothman (1976)). The sufficient cause model has evolved over the past decade to enable the detection of different forms of interaction (VanderWeele and Robins (2008); Berzuini and Dawid (2016); VanderWeele (2015); VanderWeele and Richardson (2012); Vanderweele (2010); Ramsahai (2013)).

Corresponding author: Jaffer M. Zaidi, Department of Biostatistics, Harvard University, Boston, MA 02115, USA. E-mail: jaffer.zaidi@gmail.com.

Rothman (1976) presented a model for causation as a series of different causal mechanisms, each of which is sufficient to bring about the outcome. In this model, the causal mechanisms are called "sufficient causes," defined as the minimal set of actions, events, or states of being that jointly initiate a process that will eventually result in the outcome. Many different "sufficient causes" can produce a particular outcome. For instance, in the course of treating HIV-1, viral mutations can arise. Some mutations might on their own make a particular treatment ineffective, while others might require one or more additional mutations to operate.

Within a deterministic framework, for a binary outcome, suppose we consider three known potential causes, X_1, X_2, and X_3. Suppose hypothetically that mutation X_1 and unknown factors A_1 make an individual drug resistant, denoted by binary outcome R. In contrast, mutations X_2 and X_3 are jointly sufficient if present with additional unknown factors A_2. A last mechanism might be mutations X_1 and X_2, with additional unknown factors A_3. This provides us with three sufficient causes, denoted by A_1X_1, $A_2X_2X_3$, and $A_3X_1X_2$, each of which can make a person drug resistant.

In a deterministic sufficient cause model, whenever all component causes of a particular sufficient cause are present, the outcome will definitely occur. Here, each component cause is necessary for the sufficient cause to bring about the outcome. Sufficient cause $A_2X_2X_3$ has two component causes, X_2 and X_3. This sufficient cause will not operate if either X_2 or X_3 is not present. When two component causes are both needed to cause the outcome to occur, we call this synergism. In general, it may be logically possible to represent the counterfactual outcomes across causes by different representations of sufficient causes. When every such possible sufficient cause representation has a particular conjunction, say X_2X_3, then a "sufficient cause interaction" between X_2 and X_3 is said to be present. In such cases, we then know that synergism must be present between X_2 and X_3. Scientists seek to discover synergism from data. As a result, statisticians have derived empirical conditions to enable the discovery of sufficient cause interactions (VanderWeele and Robins (2008); Berzuini and Dawid (2016); VanderWeele (2015); VanderWeele and Richardson (2012); Vanderweele (2010); Ramsahai (2013)).

In this study, we extend the sufficient cause model to categorical and ordinal outcomes, and develop the associated likelihood ratio tests. In a data application of this theory, we examine which mutations mechanistically interact in the development of HIV-1 drug resistance to Etravirine. For this applied problem, the ordinal outcome has three levels: no drug resistance, partial drug resistance, and full drug resistance.

2. Sufficient Cause Interactions for A Specified Outcome

Suppose we have an outcome Y with associated levels $Y \in \{0, 1, 2\}$. We denote binary variables $X_1, \ldots, X_k$ that each take values in $X_i \in \{0, 1\}$. Jointly, $X^k = (X_1, \ldots, X_k)$ take values within $X^k \in \{0, 1\}^k$. The individuals, denoted by ω, compose a population, denoted as Ω. We write the potential outcome $Y_{x_1,\ldots,x_k}(\omega)$ of Y for individual ω, if for $j = 1, \ldots, k$, each putative cause $X_j \in \{X_1, \ldots, X_k\}$ were set $x_j \in \{0, 1\}$. The data application considers $k = 2$. The potential outcome or counterfactual value for individual ω, had X_1 been set to x_1 and X_2 been set to x_2, is denoted as $Y_{x_1,x_2}(\omega)$. There are 3^4 potential response types, $\underline{Y}_{x_1x_2}(\omega) = (Y_{11}(\omega), Y_{10}(\omega), Y_{01}(\omega), Y_{00}(\omega))$, that form all types of individuals. We denote these as $Y_{x_1x_2}(\Omega)$, which is simply all permutations of a vector of length four, sampling with replacement from the set $\{0, 1, 2\}$.

An indicator function, denoted as $I(Y \in S)$, is used to denote a new random variable constructed from Y, which takes the value one if $Y \in S$, and zero otherwise. To construct these new binary outcomes, let $A = \{1\}$, $B = \{1, 2\}$, $C = \{2\}$, $D = \{0\}$, $E = \{0, 2\}$, and $F = \{0, 1\}$. Specifically, denote $Y^L = I(Y \in L)$, where $L \in \{A, B, C, D, E, F\}$. Potential outcome versions of Y^L are defined as $Y^L_{x_1,\ldots,x_k}(\omega) = I(Y_{x_1,\ldots,x_k}(\omega) \in L)$, where $L \in \{A, B, C, D, E, F\}$. The superscript L in the symbol Y^L does not indicate exponentiation, but rather specifies the condition used to construct this new random variable from Y. Appendix 1 provides a full list of Y^L and $Y^L_{x_1,\ldots,x_k}(\omega)$ without set notation. We require the consistency assumption, namely that $Y_{X_1(\omega),\ldots,X_k(\omega)}(\omega) = Y(\omega)$. This asummption states that the value of Y that would have been observed if $X_1, \ldots, X_k$ had been set to what they in fact were is equal to the value of Y that was observed. The consistency assumption for Y^L is implied by the consistency assumption on Y. The disjunctive operator on binary variables $X_1, \ldots, X_k$ is denoted by $\vee_{i\in\{1,\ldots k\}} X_i = X_1 \vee \cdots \vee X_k = \max\{X_1, \ldots, X_k\}$. For ease of notation, we drop the commas between the intervened variables $\{X_1, \ldots, X_k\}$ in a potential outcome; for example, $Y_{x_1,x_2}(\omega) = Y_{x_1x_2}(\omega)$.

The definitions and theorems in this section closely mimic those of VanderWeele and Robins (2008). While this paper is self-contained, Definitions 1–7 and Theorems 1–5 are logical extensions to the corresponding definitions and theorems presented in VanderWeele and Robins (2008); thus, we keep the exposition concise. Theorem 6 and Corollary 1 cannot be derived using the previous framework on sufficient causes (VanderWeele and Robins (2008, 2012); Ramsahai (2013)) based on binary outcomes.

Definition 1. (Sufficient cause for a specified outcome). We say that putative binary causes $X_1, \ldots, X_n$ are called sufficient causes for Y^L, where $L \in \{A, B, C, D, E, F\}$, if for all values of $x_1, \ldots, x_n \in X^n$ such that $x_1 \times \cdots \times x_n = 1$, we have that $Y^L_{x_1 \cdots x_n}(\omega) = 1$, for all $\omega \in \Omega' \subseteq \Omega$, where $\Omega' \neq \varnothing$.

Definition 2. (Minimal sufficient cause for a specified outcome). We say that putative binary causes $X_1, \ldots, X_n$ form a minimal sufficient cause for Y^L, where $L \in \{A, B, C, D, E, F\}$, if $X_1, \ldots, X_n$ are sufficient causes for Y^L, and no proper subset of $\{X_1, \ldots, X_n\}$ is also a sufficient cause for Y^L.

Definition 3. (Determinative sufficient causes for a specified outcome). A set of sufficient causes $M^L_1, \ldots, M^L_n$, each composed of a product of binary causes for a specified outcome Y^L, where $L \in \{A, B, C, D, E, F\}$, is defined to be determinative for Y^L if for all $\omega \in \Omega$, $Y^L_{x_1 \cdots x_s}(\omega) = 1$ if and only if $M^L_1 \vee M^L_2 \vee \cdots \vee M^L_n = 1$.

Definition 4. (Nonredundant sufficient causes for a specified outcome). A set of determinative sufficient causes $M^L_1, \ldots, M^L_n$ for Y^L, where $L \in \{A, B, C, D, E, F\}$, is called a nonredundant determinative set of minimal sufficient causes if there is no proper subset of $M^L_1, \ldots, M^L_n$ that is also a determinative set of minimal sufficient causes for Y^L.

VanderWeele and Robins (2008) note that minimality and nonredundancy should be distinguished. Minimality concerns components of a given conjunction, in that each component is necessary for the conjunction to be sufficient for the outcome to occur. Nonredundancy concerns the disjunction of conjunctions, in that each individual conjunction should be present in order for the disjunction to be determinative.

Example 1. Viral mutations can occur while an individual takes treatment for HIV. Suppose mutation X_1 enables the virus to replicate in particular cells, and mutation X_2 enables the virus to penetrate the bodies of these cells. Assume for now that these are the only two mutations that occur. A scientist may question whether mutations X_1 and X_2 are both required for this individual in order for the current treatment to become ineffective in treating HIV, which is known as drug resistance. Alternatively, would mutation X_1 on its own suffice for the individual to develop drug resistance. Scientists also grade drug resistance on an ordinal scale, partial and full. A scientist might believe that mutation X_2 alone is sufficient for partial drug resistance for a particular individual, but both X_1 and X_2 are necessary for full resistance to the same drug.

These definitions for sufficient cause for an ordinal or nominal outcome with three levels generalize the analogous definitions for a binary outcome. The defi-

nitions provided here are easily adaptable to the case where a researcher is interested in an ordinal outcome Y, where $Y \in \{0, 1, \ldots, j\}$; see the Supplementary Material. A brief exposition of this generalization is provided in Section 4. More general notions of interdependence (Ramsahai (2013)) extended to categorical and ordinal outcomes are also provided in the Supplementary Material. Denote $\bar{X}_i$ as the complement of X_i.

Theorem 1. *For putative binary causes X_1 and X_2 of specified outcome Y^L, where $L \in \{A, B, C, D, E, F\}$, there exist binary variables*

$$A_0^L(\omega), A_1^L(\omega), A_2^L(\omega), A_3^L(\omega), A_4^L(\omega), A_5^L(\omega), A_6^L(\omega), A_7^L(\omega), A_8^L(\omega),$$

which are functions of the counterfactuals $\{Y_{11}^L(\omega), Y_{10}^L(\omega), Y_{01}^L(\omega), Y_{00}^L(\omega)\}$, such that

$$\begin{aligned} Y^L = {} & A_0^L \vee A_1^L X_1 \vee A_2^L \bar{X}_1 \vee A_3^L X_2 \vee A_4^L \bar{X}_2 \vee A_5^L X_1 X_2 \\ & \vee A_6^L \bar{X}_1 X_2 \vee A_7^L X_1 \bar{X}_2 \vee A_8^L \bar{X}_1 \bar{X}_2, \end{aligned} \tag{2.1}$$

and such that

$$\begin{aligned} Y_{x_1 x_2}^L = {} & A_0^L \vee A_1^L x_1 \vee A_2^L (1 - x_1) \vee A_3^L x_2 \vee A_4^L (1 - x_2) \vee A_5^L x_1 x_2 \\ & \vee A_6^L (1 - x_1) x_2 \vee A_7^L x_1 (1 - x_2) \vee A_8^L (1 - x_1)(1 - x_2). \end{aligned}$$

The proof of Theorem 1 mimics that of the sufficient cause representation for binary outcomes provided in (VanderWeele and Robins (2008)). For completeness, we provide the proof in Supplementary Material. We call equation (2.1) a sufficient cause representation of Y^L.

We generalize our definitions and Theorem 1 to the situation where the analyst is concerned about defining and analyzing the minimum sufficient cause interaction on an ordinal variable with multiple levels, that is, more than three. This generalization is provided in the Supplementary Material. This theorem extends the results provided in VanderWeele and Robins (2008). Theorem 1 also provides a method for constructing variables A_i^L as a function of the potential outcomes that, together with disjunctions built on the set $\{X_1, X_2, \bar{X}_1, \bar{X}_2\}$, make a determinative set of sufficient causes for Y^L, where $L \in \{A, B, C, D, E\}$. Each of the conjunctions $A_0^L, A_1^L X_1, \ldots, A_8^L \bar{X}_1 \bar{X}_2$ are sufficient to cause Y^L, where $L \in \{A, B, C, D, E, F\}$. The disjunction of all of these conjunctions makes a determinative set of sufficient causes for Y^L, where $L \in \{A, B, C, D, E, F\}$. Similarly to the binary outcome context, A_i^L variables could be considered as unknown factors that, together with the associated conjunction of $\varnothing, X_1, X_2, \bar{X}_1,$

$\bar{X}_2, X_1\bar{X}_2, \bar{X}_1X_2$, and $\bar{X}_1\bar{X}_2$, complete the sufficient cause for specified outcome Y^L, where $L \in \{A, B, C, D, E, F\}$.

Now that we have defined sufficient cause for a specified outcome, we define sufficient cause interactions for a specified outcome. Based upon these definitions, counterfactual and empirical conditions are derived to detect the presence of sufficient cause interactions for a specified outcome.

Example 2. Consider the drug resistance example presented earlier. Suppose we have two types of individuals in our population. Individual 1 develops full drug resistance if she has either of the two mutations, while individual two develops full drug resistance only if she has both mutations. The construction of the variables A_i from Theorem 1 yields

$$Y^C(\omega) = A_1^C(\omega)X_1(\omega) \vee A_2^C(\omega)X_2(\omega) \vee A_8^C(\omega)X_1(\omega)X_2(\omega),$$

where Y^C denotes full drug resistance. Suppose these same two individuals would develop partial drug resistance if they have either of the two mutations. Then, applying of Theorem 1 yields

$$Y^A(\omega) = X_1(\omega) \vee X_2(\omega),$$

where $Y^A(\omega)$ denotes partial drug resistance.

Definition 5. (Minimal sufficient cause interaction for a specified outcome). Suppose $F_1 \in \{X_1, \bar{X}_1\}$ and $F_2 \in \{X_2, \bar{X}_2\}$. If in every nonredundant minimal sufficient cause representation for a specified outcome Y^L, where $L \in \{A, B, C, D, E, F\}$, we are able to find a sufficient cause that contains F_1F_2, then we say that the conjunction F_1F_2 exhibits or displays minimal sufficient cause interaction for outcome Y^L.

Definition 6. (Irreducible sufficient cause interactions for a specified outcome). Suppose $F_1 \in \{X_1, \bar{X}_1\}$ and $F_2 \in \{X_2, \bar{X}_2\}$. If in every sufficient cause representation for Y^L, where $L \in \{A, B, C, D, E, F\}$, we are able to find a sufficient cause that contains F_1F_2, then F_1F_2 is said to be irreducible for Y^L.

These two definitions are shown to be equivalent in our case: that is, an irreducible sufficient cause interaction for a specified outcome Y^L is a minimal sufficient cause interaction for the same outcome, and vice versa. The theorem and proof demonstrating that the definitions are equivalent replicates the arguments of VanderWeele and Robins (2008), and as such are omitted. Here, we say that the effects of $F_1 \in \{X_1, \bar{X}_1\}$ and $F_2 \in \{X_2, \bar{X}_2\}$ on a specified outcome Y^L,

where $L \in \{A, B, C, D, E, F\}$, are synergistic, or represent synergism, if there is a sufficient cause for Y^L, such that F_1F_2 is contained within its conjunction. The rest of the proofs of the theorems and corollaries are collected in the Supplement Material.

Theorem 2. *Suppose $L \in \{A, B, C, D, E, F\}$. There exists an individual $\omega \in \Omega$ for whom $Y^L_{11}(\omega) = 1$ and $Y^L_{10}(\omega) = Y^L_{01}(\omega) = 0$, if and only if the conjunction X_1X_2 exhibits a sufficient cause interaction for a specified outcome Y^L.*

We now consider empirical conditions for detecting sufficient cause interactions. The symbol $\amalg$ is used to denote independence. For example, $Y \amalg X_1$ denotes that Y is marginally independent of X_1, and $Y \amalg X_1 \mid X_2$ denotes that Y is conditionally independent of X_1, given X_2.

Theorem 3. *Suppose V is a set of variables that are sufficient to control for the confounding of X_1 and X_2 on Y^L, where $L \in \{A, B, C, D, E, F\}$, that is, $Y^L_{x_1x_2} \amalg \{X_1, X_2\} \mid V$. We conclude that X_1X_2 exhibit sufficient cause interaction for a specified outcome Y^L if for some value v of V, the following inequality holds:*

$$\begin{aligned} 0 < {} & E(Y^L \mid X_1 = 1, X_2 = 1, V = v) \\ & - E(Y^L \mid X_1 = 1, X_2 = 0, V = v) \\ & - E(Y^L \mid X_1 = 0, X_2 = 1, V = v). \end{aligned} \tag{2.2}$$

Henceforth, we use the shorthand notation $p^L_{x_1x_2}$ to denote $P(Y \in L \mid X_1 = x_1, X_2 = x_2)$ and $p^L_{x_1x_2v}$ to denote $P(Y \in L \mid X_1 = x_1, X_2 = x_2, V = v)$. Replacing X_1 or X_2 by either or both of their complements yields similar results for antagonism. The results in Theorem 3 generalize those for identifying synergism for a binary outcome, as established in VanderWeele and Richardson (2012), to categorical or ordinal outcomes under a specified condition. Theorems 2 and 3 have generalizations for categorical or ordinal outcomes with an arbitrary number of levels; see Section 5. Our approach allows the researcher to detect a sufficient cause interaction between two variables for an ordinal outcome under specified conditions at different levels, or for an amalgam of different levels of the categorical or ordinal outcome. We provide an example to illustrate Theorem 3.

Example 3. Consider $Y^C = I(Y = 2)$. Now, consider the left-hand side of the inequality (2.2),

$$\begin{aligned} & E(Y^C \mid X_1 = 1, X_2 = 1, V = v) - E(Y^C \mid X_1 = 1, X_2 = 0, V = v) \\ & \quad - E(Y^C \mid X_1 = 0, X_2 = 1, V = v) \end{aligned}$$

$$
\begin{aligned}
&= E(I(Y=2) \mid X_1=1, X_2=1, V=v) - E(I(Y=2) \mid X_1=1, X_2=0, V=v) \\
&\quad - E(I(Y=2) \mid X_1=0, X_2=1, V=v) \\
&= P(Y=2 \mid X_1=1, X_2=1, V=v) - P(Y=2 \mid X_1=1, X_2=0, V=v) \\
&\quad - P(Y=2 \mid X_1=0, X_2=1, V=v) \\
&= p^C_{11v} - p^C_{10v} - p^C_{01v}.
\end{aligned}
$$

Therefore, if $p^C_{11v} - p^C_{10v} - p^C_{01v} > 0$, we can say that X_1X_2 exhibits sufficient cause interaction for the outcome $I(Y=2)$, or equivalently, that X_1X_2 exhibits sufficient cause interaction for the ordinal outcome Y at level 2.

Following the same steps as in Example 1, we can show that if $p^C_{11v} + p^A_{11} - p^C_{10v} - p^A_{10v} - p^C_{01v} - p^A_{01v} > 0$, X_1X_2 exhibits sufficient cause interaction for the outcome $I(Y \geq 1)$, or equivalently, that X_1X_2 exhibits sufficient cause interaction for the ordinal outcome Y at the level 1 or 2. Similarly, if $p^D_{11v} + p^C_{11} - p^D_{10v} - p^C_{10v} - p^D_{01v} - p^C_{01v} > 0$, X_1X_2 exhibits sufficient cause interaction for the outcome $I(Y \in \{0,2\})$.

VanderWeele and Robins (2008) demonstrate that if it can be assumed that variables have positive monotonic effects on a binary outcome (that is, the variables never prevent the outcome), then one can use less stringent tests to detect sufficient cause interaction than if one were unable to make this assumption. We now examine the analogous results in the case of an ordinal outcome with three levels. If Y were categorical, the same definitions and results, namely Definitions 1–6 and Theorems 1–3, hold. Results that require monotonicity work only with ordinal outcomes, because the next definition requires the outcome to be ordinal. Therefore, Theorems 4–6 and Corollary 1 are only valid for ordinal outcomes, and cannot be applied to categorical outcomes.

Definition 7. (Monotonic effect for an ordinal outcome with three levels). For any two binary variables X_1 and X_2, if for all $\omega \in \Omega$, $Y_{x_1x_2}(\omega)$ is nondecreasing in x_1 for any given $x_2 \in X$, then we say that X_1 has a positive monotonic effect on Y. Similarly, if for all $\omega \in \Omega$, $Y_{x_1x_2}(\omega)$ is nondecreasing in x_2 for any given $x_1 \in X$, then we say that X_2 has a positive monotonic effect on Y.

Theorem 4. *Suppose X_1 and X_2 both have positive monotonic effects on ordinal variable Y, and that $Y^B_{x_1x_2} \amalg \{X_1, X_2\} \mid V$. If, for some value $v \in V$, we have*

$$p^B_{11v} - p^B_{10v} - p^B_{01v} + p^B_{00v} > 0,$$

then X_1 and X_2 display synergism for outcome $Y^B = I(Y \geq 1)$.

Theorem 5. *Suppose X_1 and X_2 both have positive monotonic effects on ordinal variable Y, and that $Y^C_{x_1x_2} \amalg \{X_1, X_2\} \mid V$. If, for some value $v \in V$, we have*

$$p^C_{11v} - p^C_{10v} - p^C_{01v} + p^C_{00v} > 0,$$

then X_1 and X_2 display synergism for outcome $Y^C = I(Y = 2)$.

These results are the same as those that would have been established had we dichotomized the outcome at the outset and applied the empirical conditions established from VanderWeele and Robins (2008). The next theorem provides a result that cannot be derived from existing literature on sufficient cause interaction or mechanistic interaction (VanderWeele and Robins (2008); Berzuini and Dawid (2016); Ramsahai (2013); VanderWeele and Richardson (2012)). This is a novel result that enables researchers to discover a sufficient cause interaction for a specified outcome $Y^A = I(Y = 1)$. The proofs of these results are collected in the Supplementary Material.

Theorem 6. *Suppose X_1 and X_2 both have positive monotonic effects on ordinal variable Y, and that $Y^A_{x_1x_2} \amalg \{X_1, X_2\} \mid V$. If, for some value $v \in V$, we have at least one of the following inequalities:*

$$p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{01v} > 0, \tag{2.3}$$
$$p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{10v} > 0, \tag{2.4}$$
$$p^A_{11v} - p^A_{10v} - p^A_{01v} > 0, \tag{2.5}$$

then X_1 and X_2 display synergism for outcome $Y^A = I(Y = 1)$.

Corollary 1. *Suppose X_1 and X_2 both have positive monotonic effects on ordinal variable Y, and that $Y^A_{x_1x_2} \amalg \{X_1, X_2\} \mid V$. If, for some value $v \in V$, we have at least one of the following inequalities:*

$$2 \cdot p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} > 1, \tag{2.6}$$
$$p^A_{11v} - p^D_{11v} - p^C_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} > 0 \tag{2.7}$$
$$2p^A_{11v} - p^A_{10v} - p^A_{01v} - p^D_{00v} - p^C_{00v} > 0 \tag{2.8}$$
$$p^A_{11v} - p^A_{10v} - p^A_{01v} > 0, \tag{2.9}$$

then X_1 and X_2 display synergism for outcome $Y^A = I(Y = 1)$.

We demonstrate in the Supplementary Material that if $2 \cdot p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} > 1$, then $p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{01v} > 0$ and $p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{01v} > 0$. The converse is not true. This implies that the

empirical conditions (2.3) and (2.4) are weaker than the empirical condition (2.6). Conditions (2.6)–(2.8) are shown to be equivalent to one another in the Supplementary Material. The only circumstance in which we would use condition (2.6) instead of conditions (2.3) and (2.4) is when we do not have data on the outcome $I(Y = 2)$. Note that Theorems 4–6 are derived in the Supplementary Material using arguments made from the sufficient cause framework and monotonicity. We derived the same inequality constraints using a different approach that modifies the theory provided in Ramsahai (2013) using convex polytopes from binary outcomes to categorical and ordinal outcomes. The results show that the empirical conditions presented here are the only inequalities observed. A more detailed explanation of how the Ramsahai approach is adapted to the ordinal outcome setting is provided in the Supplementary Material.

2.1. Inference for sufficient cause interaction for ordinal outcomes

Previous authors have used likelihood ratio tests to conduct hypothesis tests on moment conditions that stem from problems in causal inference (Ramsahai (2013); Ramsahai and Lauritzen (2011)), including the sufficient cause interaction for binary outcomes (Ramsahai (2013)). The saturated Bernoulli model has been proposed to detect sufficient cause interactions for binary outcomes in the presence of covariates (VanderWeele and Richardson (2012); VanderWeele and Robins (2008); Vansteelandt, VanderWeele and Robins (2012)). Researchers have also used Bonferonni corrections to test multiple moment conditions in the causal inference literature (Wang, Robins and Richardson (2017)). The approach taken here uses likelihood ratio tests (Ramsahai and Lauritzen (2011); Ramsahai (2013)). In the setting of a composite null, the likelihood ratio test statistics' asymptotic distribution is obtained assuming the true parameter is on the boundary of the null hypothesis (Van der Vaart (2000); Ramsahai (2013); Ramsahai and Lauritzen (2011); Drton (2009)). For the tests considered in the data analysis, the likelihood ratio test follows a weighted mixture of χ^2-distributions (Ramsahai (2013)). A description of the asymptotics of likelihood ratio tests under multiple inequality constraints is available in Silvapulle and Sen (2011). Likelihood ratio tests that use Theorems 3–5 are closely related to those under inequality constraints for sufficient cause interaction (Ramsahai (2013)). Theorems 3–6 provide the alternative space to each of the specified forms of sufficient cause interaction. As usual, the complement of the alternative space is the null space.

Theorems 3–5 only involve a statistical test with a single inequality constraint. For example, the hypothesis test $H_0 : p^A_{11} - p^A_{10} - p^A_{01} \leq 0$ versus $H_1 : p^A_{11} - p^A_{10} - p^A_{01} > 0$ is a test to establish whether X_1 and X_2 display syn-

ergism for outcome $Y^A = I(Y = 1)$. For such hypothesis tests that involve a single inequality constraint, the null space is a half-space (Ramsahai and Lauritzen (2011); Ramsahai (2013); Self and Liang (1987)). Throughout, we let t denote the observed value of the likelihood ratio statistic. For these tests with a single inequality constraint, following Self and Liang, the p-value of the likelihood ratio test is $P(\chi_1^2 > t)/2$ for positive t, and one otherwise (Self and Liang (1987); Ramsahai (2013)). On the other hand, for Theorem 6, the null space is defined by the intersection of three half-spaces, each of which is defined using an inequality constraint. To use Theorem 6, the associated null space is the intersection of the following three inequalities:

$$\begin{aligned} p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{01v} &\leq 0, \\ p^A_{11v} - p^A_{10v} - p^A_{01v} + p^A_{00v} + p^C_{00v} - p^C_{10v} &\leq 0, \\ p^A_{11v} - p^A_{10v} - p^A_{01v} &\leq 0. \end{aligned}$$

A similar situation arises in the falsification of the binary instrumental variable model (Ramsahai and Lauritzen (2011)). The correct p-value depends on where the true parameter lies on the boundary of the null space (Ramsahai and Lauritzen (2011)). If the true parameter lies on the boundary of only one of the inequality constraints, then the correct p-value is $P(\chi_1^2 > t)/2$ (Ramsahai and Lauritzen (2011)). The correct asymptotic sampling distribution is clear if the true parameter lies on the boundary of multiple inequality constraints in the context of the falsification of binary instrumental variable model (Ramsahai and Lauritzen (2011)). However, the statistician will not always be able to easily derive the associated asymptotic sampling distribution when there is more than one inequality constraint that defines the null space. This is particularly true when there is a large number of inequality constraints.

For our data analysis, the only instance in which we examine a null space defined through multiple inequality constraints stems from Theorem 6. For these types of hypothesis test, the asymptotic sampling distribution changes, depending on where we assume the true parameter lies on the null space. To construct a p-value based on Theorem 6, we use Theorem 3 of Self and Liang (1987), finding that the asymptotic sampling distribution is $w_{0,3}\chi_0^2 + w_{1,3}\chi_1^2 + w_{2,3}\chi_2^2 + w_{3,3}\chi_3^2$ if the true parameter lies on the boundary of all three half-spaces. The weights can be calculated using equations (4.8) and (4.9) in Shapiro (1985). If the true parameter lies the boundary of on two of the three half-spaces, then the asymptotic sampling distribution is given by $w_{0,2}\chi_0^2 + w_{1,2}\chi_1^2 + w_{2,2}\chi_2^2$. Finally, if the true parameter lies on the boundary of only one of the three half-spaces, then we can use the earlier

Table 1. Drug Resistance to Etravirine by mutation category

$X_1=0, X_2=0$			$X_1=0, X_2=1$			$X_1=1, X_2=0$			$X_1=1, X_2=1$		
$n_{0,0}$	P	F	$n_{0,1}$	P	F	$n_{1,0}$	P	F	$n_{1,1}$	P	F
445	74	57	17	2	1	10	2	1	12	7	4

Here, $n_{x_1,x_2} = \sum_{i=1}^{n} I(X_{1i} = x_1, X_{2i} = x_2)$. P denotes individuals with partial drug resistance, and F denotes full drug resistance.

p-value of $P(\chi_1^2 > t)/2$. To get one p-value, one can use the least favorable configuration p-value $= \sup_{\mathbf{p}\in\mathbf{p}_0} P(T > t)$, where T is the likelihood-ratio test statistic defined in the Supplementary Material, t is the observed test statistic, $\mathbf{p}$ is the parameter space defined in the Supplementary Material, and $\mathbf{p}_0$ is the null parameter space defined by the relevant inequality constraints. The statistics and econometrics literature for calculating the asymptotic sampling distribution of test statistics in the analysis of moment structures under inequality constraints is vast (Drton (2009); Geyer (1994); Shapiro (1985); Wolak (1991); Dardanoni and Forcina (1998); Silvapulle and Sen (2011)). When analytic formulae to calculate the weights of the χ^2 distributions are not available, Monte Carlo methods can be used to determine the weights to a prespecified degree of precision (Dardanoni and Forcina (1998); Silvapulle and Sen (2011)). The likelihood and likelihood ratio test statistic are provided in the Supplementary Material.

3. Application to HIV Drug Resistance

HIV drug resistance arises when viral mutations make particular drugs less effective in controlling HIV (Tang and Shafer (2012)). For our data analysis, we consider data from the Stanford HIV drug resistance database on 484 viral isolates, summarized in Table 1, which presents information on viral resistance to the NRTI Etravirine (Tang and Shafer (2012)). The two viral mutations under consideration are X_1, denoting the presence of mutation 103 R, and X_2, denoting the presence of mutation 179 D. The scientific question under consideration is whether mutations 103 R and 179 D interact synergystically to confer drug resistance to Etravirine. Our outcome Y is an ordinal outcome with three levels: no drug resistance, partial drug resistance, and full drug resistance. Here, no drug resistance is labeled zero, partial drug resistance is labeled one, and full drug resistance is labeled two. We assume there is no confounding between Y and X_1, X_2. This means that $Y_{x_1x_2}$ II (X_1, X_2). A laboratory experiment provides a contingency table, summarized in Table 1, on HIV drug resistance by mutation category (Tang and Shafer (2012)).

Table 2. Likelihood ratio test of drug resistance

Outcome	Null Hypothesis	LRT	p-value
$I(Y=1)$	$p^A_{11}-p^A_{10}-p^A_{01}\leq 0$	1.477	0.112
$I(Y\geq 1)$	$p^B_{11}-p^B_{10}-p^B_{01}\leq 0$	4.218	0.020
$I(Y=2)$	$p^C_{11}-p^C_{10}-p^C_{01}\leq 0$	0.925	0.168

Table 3. Likelihood ratio test of drug resistance under monotonicity

Outcome	Null Hypothesis	LRT	p-value
	$p^A_{11}-p^A_{10}-p^A_{01}\leq 0$		
$I(Y=1)$	$p^A_{11}-p^A_{10}-p^A_{01}+p^A_{00}+p^C_{00}-p^C_{01}\leq 0$	4.704	$0.057^\star$
	$p^A_{11}-p^A_{10}-p^A_{01}+p^A_{00}+p^C_{00}-p^C_{10}\leq 0$		
$I(Y\geq 1)$	$p^B_{11}-p^B_{10}-p^B_{01}+p^B_{00}\leq 0$	10.624	< 0.005
$I(Y=2)$	$p^C_{11}-p^C_{10}-p^C_{01}+p^C_{00}\leq 0$	2.585	0.054

The p-value annotated with a $\star$ is obtained under the least favorable configuration.

Table 2 provides the likelihood ratio test statistics and associated p-values for assessing the sufficient cause interaction for each of the specified outcome levels $I(Y=1)$, $I(Y\geq 1)$, and $I(Y=2)$. Table 3 provides the same information, but assumes that the effects of X_1 and X_2 are positive monotonic for Y. If one is unwilling to make any monotonicity assumptions, there is no statistical evidence that X_1 and X_2 have a synergistic effect on $Y^C=I(Y=2)$ or $Y^A=I(Y=1)$, but there is some evidence for a synergistic effect on $Y^B=I(Y\geq 1)$. If one is willing to assume that X_1 and X_2 have positive monotonic effects on the outcome, the evidence for a synergistic effect of X_1 and X_2 on specified outcomes $I(Y=1)$ and $I(Y=2)$ is stronger, although the p-values are slightly above the nominal 0.05 rejection threshold. In this situation, we have evidence that X_1 and X_2 have a synergistic effect on $I(Y\geq 1)$, and thus that X_1 and X_2 have a synergistic effect on either or both of the outcomes $I(Y=1)$ and $I(Y=2)$, because for $p^B_{11}-p^B_{10}-p^B_{01}>0$, either or both of these inequalities $p^C_{11}-p^C_{10}-p^C_{01}>0$ or $p^A_{11}-p^A_{10}-p^A_{01}>0$ needs to hold. However, given the current sample size, we are unable to detect whether that synergistic effect occurs either for $I(Y=1)$ or $I(Y=2)$.

Discussions on whether the proposed exposures have monotonic effects on the outcome should occur with the scientific investigators. Such assumptions could have scientific justification, and would enable researchers to use less stringent conditions to draw the same inferences. Primary mutations, such as the two investigated here, "directly decrease the susceptibility of the virus to an an-

tiretroviral treatment"(Tang and Shafer (2012)). To the best of our knowledge, it is not known whether mutations 103 R and 179 D are never preventative for partial or full drug resistance for every individual taking Etravirine as part of their treatment for HIV. Monotonicity assumptions can be falsified from the data, but they are never completely verifiable.

4. Generalizations and Extensions

In this section, we allow our ordinal outcome Y to take values $Y \in \{0, 1, \ldots, j\}$. If we want to investigate whether putative binary causes X_1 and X_2 have synergistic effects on an outcome $Y^y = I(Y \geq y)$, where $0 < y \leq j$, then assuming no confounding between putative causes X_1 and X_2 on Y^y, we need to check that $P(Y \geq y \mid X_1 = 1, X_2 = 1) - P(Y \geq y \mid X_1 = 1, X_2 = 0) - P(Y \geq y \mid X_1 = 0, X_2 = 1) > 0$. The proof of this result is similar to the proof of Theorem 3, and is provided in the Supplementary Material.

Theorem 7. *If we can assume that X_1 and X_2 have positive monotonic effects on Y, then $P(Y \geq y \mid X_1 = 1, X_2 = 1) - P(Y \geq y \mid X_1 = 1, X_2 = 0) - P(Y \geq y \mid X_1 = 0, X_2 = 1) + P(Y \geq y \mid X_1 = 0, X_2 = 0) > 0$ implies X_1 and X_2 display synergism for $I(Y \geq 1)$.*

Let $S \subseteq \{1, 2, \ldots, n-1\}$, $Y^S = I(Y \in S)$, $S^+ = \max(S)$, and $Y^{S^+} = I(Y > S^+)$. If we wish to check whether X_1 and X_2 have synergistic effects on $Y^S = I(Y \in S)$, where S is an arithmetic progression with common difference one, and X_1 and X_2 both have positive monotonic effects on Y, then we need to check whether at least one of the following three inequalities hold:

$$
\begin{aligned}
&P(Y \in S \mid X_1 = 1, X_2 = 1) \\
&-P(Y \in S \mid X_1 = 1, X_2 = 0) - P(Y \in S \mid X_1 = 0, X_2 = 1) > 0, \\
&P(Y \in S \mid X_1 = 1, X_2 = 1) - P(Y \in S \mid X_1 = 1, X_2 = 0) \\
&-P(Y \in S \mid X_1 = 0, X_2 = 1) + P(Y \in S \mid X_1 = 0, X_2 = 0) \\
&+P(Y > S^+ \mid X_1 = 0, X_2 = 0) - P(Y > S^+ \mid X_1 = 1, X_2 = 0) > 0, \\
&P(Y \in S \mid X_1 = 1, X_2 = 1) - P(Y \in S \mid X_1 = 1, X_2 = 0) \\
&-P(Y \in S \mid X_1 = 0, X_2 = 1) + P(Y \in S \mid X_1 = 0, X_2 = 0) \\
&+P(Y > S^+ \mid X_1 = 0, X_2 = 0) - P(Y > S^+ \mid X_1 = 0, X_2 = 1) > 0.
\end{aligned}
$$

Here, the second and third inequalities are implied by $P(Y \in S \mid X_1 = 1, X_2 = 1) - P(Y \notin S \mid X_1 = 1, X_2 = 1) - P(Y \in S \mid X_1 = 1, X_2 = 0) - P(Y \in S \mid X_1 = 0, X_2 = 1) + P(Y \in S \mid X_1 = 0, X_2 = 0) > 0$. This last inequality

can be shown to be equivalent to $2 \cdot P(Y \in S \mid X_1 = 1, X_2 = 1) - P(Y \in S \mid X_1 = 1, X_2 = 0) - P(Y \in S \mid X_1 = 0, X_2 = 1) + P(Y \in S \mid X_1 = 0, X_2 = 0) > 1$. If X_1 or X_2 do not have positive monotonic effects on Y, or S is not an arithmetic progression with common difference one, then only the first of the three inequalities remains valid for determining whether X_1 and X_2 have synergistic effects on $Y^S = I(Y \in S)$. The full set of definitions, theorems, and proofs associated with this generalization are provided in the Supplementary Material.

For this section, we allow $(Y_{11}, Y_{10}, Y_{01}, Y_{00})$ to have a distribution function $P(Y_{11} \in y_{11}, Y_{10} \in y_{10}, Y_{01} \in y_{01}, Y_{00} \in y_{00})$, where $y_{11}, y_{10}, y_{01}, y_{00}$ are all subsets of $\mathbf{R}$. The proofs of next two results are provided in the Supplementary Material.

Definition 8. (Generalized Positive Monotonicity). We say that X_1 has a positive monotonic effect on $Y \in y_c$, for any fixed $y_c \subset \mathbf{R}$, if there is no individual $\omega \in \Omega$ such that $Y_{x_1x_2}(\omega) \notin y_c$ and $Y_{x_3x_2}(\omega) \in y_c$, for all $x_1 > x_3$ and any fixed x_2. Similarly, we say that X_2 has a positive monotonic effect on $Y \in y_c$, for some $y_c \subset \mathbf{R}$, if there is no individual $\omega \in \Omega$ such that $Y_{x_1x_2}(\omega) \notin y_c$ and $Y_{x_1x_3}(\omega) \in y_c$, for all $x_2 > x_3$ and any fixed x_1. If X_1 and X_2 each individually have a positive monotonic effect on $Y \in y_c$, for any fixed $y_c \subset \mathbf{R}$, then we say that X_1 and X_2 have positive monotonic effects on $Y \in y_c$.

Theorem 8. *Suppose* $Y_{x_1x_2} \amalg X_1X_2$. *Here,* y_a *is any subset of* $\mathbf{R}$. *The contrast*

$$P(Y \in y_a \mid X_1 = 1, X_2 = 1) - P(Y \in y_a \mid X_1 = 1, X_2 = 0) \\ -P(Y \in y_a \mid X_1 = 0, X_2 = 1)$$

is equal to

$$P(Y_{11} \in y_a, Y_{10} \notin y_a, Y_{01} \notin y_a) - P(Y_{11} \notin y_a, Y_{10} \in y_a, Y_{01} \notin y_a) \\ -P(Y_{11} \in y_a, Y_{10} \in y_a, Y_{01} \in y_a) - P(Y_{11} \notin y_a, Y_{10} \notin y_a, Y_{01} \in y_a).$$

Theorem 9. *Suppose* $Y_{x_1x_2} \amalg X_1X_2$, *and that* X_1 *and* X_2 *have positive monotonic effects on* $Y \in y_c$. *For any* y_c *that is a subset of* $\mathbf{R}$, *the contrast*

$$P(Y \in y_c \mid X_1 = 1, X_2 = 1) - P(Y \in y_c \mid X_1 = 1, X_2 = 0) \\ -P(Y \in y_c \mid X_1 = 0, X_2 = 1) + P(Y \in y_c \mid X_1 = 0, X_2 = 0)$$

is equal to $P(Y_{11} \in y_c, Y_{10} \notin y_c, Y_{01} \notin y_c, Y_{00} \notin y_c)$.

5. Discussion

Our extension of the sufficient cause model to ordinal and categorical outcomes enables researchers to investigate more complex scientific questions. In addition, we derive novel empirical conditions that in some situations are more powerful in testing the sufficient cause interaction for ordinal outcomes than applying previously formulated empirical conditions for a binary outcome to a dichotomized ordinal outcome. The interpretations of sufficient cause interaction are far stronger than those of tests for statistical interaction. We applied these novel tests to detect whether the viral mutations 103R and 179D interact synergystically to confer partial, full, or any drug resistance to Etravirine.

Supplementary Material

The online Supplementary Material provides definitions and proofs of the theorems and corollaries presented here.

Acknowledgments

The authors acknowledge the helpful discussions and correspondence with Denis Agniel, Ron Bosch, Mathias Drton, James M. Robins, Robert W. Shafer, Eric J. Tchetgen Tchetgen, and Linbo Wang. This research was supported by the National Institutes of Health, United States.

Appendix

A. Notation

We denote $Y^A = I(Y \in \{1\})$, $Y^B = I(Y \in \{1,2\})$, $Y^C = I(Y \in \{2\})$, $Y^D = I(Y \in \{0\})$, $Y^E = I(Y \in \{0,2\})$, $Y^F = I(Y \in \{0,2\})$. Potential outcome versions of Y^A, Y^B, Y^C, Y^D, Y^E, and Y^F are defined as $Y^A_{x_1,\ldots,x_k}(\omega) = I(Y_{x_1,\ldots,x_k}(\omega) \in \{1\})$, $Y^B_{x_1,\ldots,x_k} = I(Y_{x_1,\ldots,x_k}(\omega) \in \{1,2\})$, $Y^C_{x_1,\ldots,x_k} = I(Y_{x_1,\ldots,x_s}(\omega) \in \{2\})$, $Y^D_{x_1,\ldots,x_s} = I(Y_{x_1,\ldots,x_k}(\omega) \in \{0\})$, $Y^E_{x_1,\ldots,x_s} = I(Y_{x_1,\ldots,x_k}(\omega) \in \{0,2\})$, and $Y^F_{x_1,\ldots,x_s} = I(Y_{x_1,\ldots,x_k}(\omega) \in \{0,1\})$.

References

Berzuini, C. and Dawid, A. P. (2016). Stochastic mechanistic interaction. *Biometrika* **103**, 89–102.

Cayley, A. (1853). XXXVII. Note on a question in the theory of probabilities. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **6**, 259–259.

Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochas-

tic orderings in a nonparametric context. *Journal of the American Statistical Association* **93**, 1112–1123.

Drton, M. (2009). Likelihood ratio tests and singularities. *The Annals of Statistics* **37**, 979–1012.

Geyer, C. J. (1994). On the asymptotics of constrained m-estimation. *The Annals of Statistics* **22**, 1993–2010.

Ramsahai, R. and Lauritzen, S. (2011). Likelihood analysis of the binary instrumental variable model. *Biometrika* **98**, 987–994.

Ramsahai, R. R. (2013). Probabilistic causality and detecting collections of interdependence patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 705–723.

Rothman, K. J. (1976). Causes. *American Journal of Epidemiology* **104**, 587–592.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72**, 133–144.

Silvapulle, M. J. and Sen, P. K. (2011). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. John Wiley & Sons.

Tang, M. W. and Shafer, R. W. (2012). HIV-1 antiretroviral resistance. *Drugs* **72**, e1–e25.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Vanderweele, T. J. (2010). Sufficient cause interactions for categorical and ordinal exposures with three levels. *Biometrika* **97**, 647–659.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

VanderWeele, T. J. and Richardson, T. S. (2012). General theory for interactions in sufficient cause models with dichotomous exposures. *Annals of statistics* **40**, 2128–2161.

VanderWeele, T. J. and Robins, J. M. (2008). Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika* **95**, 49–61.

VanderWeele, T. J. and Robins, J. M. (2012). Stochastic counterfactuals and stochastic sufficient causes. *Statistica Sinica* **22**, 379–392.

Vansteelandt, S., VanderWeele, T. J. and Robins, J. M. (2012). Semiparametric tests for sufficient cause interaction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 223–244.

Wang, L., Robins, J. M. and Richardson, T. S. (2017). On falsification of the binary instrumental variable model. *Biometrika* **104**, 229–236.

Wolak, F. A. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica: Journal of the Econometric Society* **59**, 981–995.

Jaffer M. Zaidi

Department of Biostatistics, Harvard University, 677 Huntington Avenue, Kresge Building, Boston, MA 02115, USA.

E-mail: jaffer.zaidi@gmail.com

Tyler J. VanderWeele

Department of Epidemiology, Department of Biostatistics, Harvard University, 677 Huntington Avenue, Kresge Building, Boston, MA 02115, USA.

E-mail: tvanderw@hsph.harvard.edu

(Received June 2019; accepted June 2020)