# Deep determinism and the assessment of mechanistic interaction between categorical and continuous variables

CARLO BERZUINI and A. PHILIP DAWID

Statistical Laboratory,
Centre for Mathematical Sciences,
University of Cambridge, UK

[c.berzuini,apd]@statslab.cam.ac.uk

**Summary**

Our aim is to detect mechanistic interaction between the effects of two causal factors on a binary response, as an aid to identifying situations where the effects are mediated by a common mechanism. We propose a formalization of mechanistic interaction which acknowledges asymmetries of the kind "factor $A$ interferes with factor $B$, but not viceversa". A class of tests for mechanistic interaction is proposed, which works on discrete or continuous causal variables, in any combination. Conditions under which these tests can be applied under a generic regime of data collection, be it interventional or observational, are discussed in terms of conditional independence assumptions within the framework of Augmented Directed Graphs. The scientific relevance of the method and the practicality of the graphical framework are illustrated with the aid of two studies in coronary artery disease. Our analysis relies on the "deep determinism" assumption that there exists some relevant set $V$ — possibly unobserved — of "context variables", such that the response $Y$ is a deterministic function of the values of $V$ and of the causal factors of interest. Caveats regarding this assumption in real studies are discussed.

## 1 Introduction

Let the binary random variable $Y$ indicate occurrence ($Y$=1) or non-occurrence ($Y$=0) of an outcome event of interest, and let $Y$ depend causally (in a sense to be later clarified) on factors $A$ and $B$. Also consider a real but possibly

unobservable variable or set of variables $V$, which collude with $A$ and $B$ to cause the response $Y$, as illustrated by the directed graph of Figure 1*(a)*. In general, even were we to know $A$, $B$ and $V$, the response $Y$ would not be fully determined, but would retain an element of random variation. In certain applications, however, it might be reasonable to assume that there exists some relevant set of variables $V$, which we will term *context variables*, such that the binary response $Y$ is fully determined, without further variation, by $V$ and the values we impose on $A$ and $B$. More precisely, consider the collection of (real or hypothetical) interventional regimes where we force $A$ and $B$ to take on some configuration $(a, b)$. Then the assumption is that, under such regimes, we have:

$$Y = f(A, B, V) \tag{1}$$

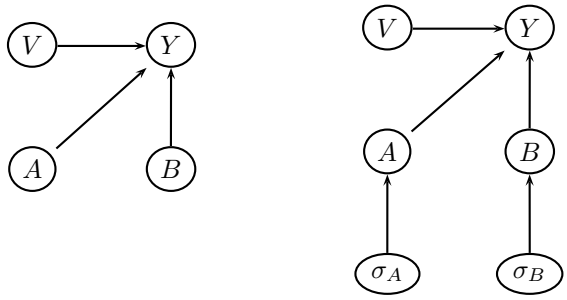
for some (typically unknown) function $f$. Thus, for any value of $V$, the $(a, b)$ configuration which we force upon $(A, B)$ will precisely dictate whether or not the event $Y = 1$ will occur. We call this assumption *deep determinism*.

If we can perform an experiment, setting $A$ and $B$ to specific values and observing the corresponding $Y$ outcomes (but not observing $V$), the resulting data may help us predict the effect upon $Y$ of intervening on $A$ and/or $B$. But we can probe more deeply. We can investigate *context-specific* causal effects — the effects of $A$ and $B$ upon $Y$ in a context determined by some given value $v$ for $V$. For example, if $A$ and $B$ are logical variables, then for any fixed value $v$ of $V$ the $f$ function of Equation (1) will take one of sixteen possible Boolean patterns, such as, for example, $Y_{=}A \vee B$, or $Y_{=}\overline{A} \wedge B$, and so on. Under appropriate assumptions, the researcher may be able to infer that a certain pattern occurs in a random individual with positive probability. If the pattern is, say, $Y_{=}A \wedge B$ — a pattern where the two effects are interdependent — one might take this as evidence that, in certain circumstances, $A$ and $B$ operate in the same mechanism. [12], [13], [18], [16], [19] and [15] have explored this territory, and proposed a series of empirical conditions for "interdependence" of binary variables focused on mechanistic interaction. [17] extends this theory to multi-level ordered categorical factors.

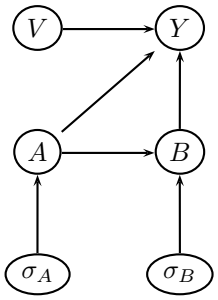
The mathematical form of the tests proposed here is similar to those that the above authors have proposed for discrete causal factors. However, by introducing novel assumptions, we derive tests valid in the more general case of categorical and continuous causal factors, in any combination.

We also provide a different justification and different assumptions for inference about mechanism, in a framework built around the above notion of deep determinism.

Section 2 introduces the concept of *interference* to capture the idea of two variables, $A$ and $B$, influencing $Y$ by operating through the same mechanism; this concept allows for asymmetry in the way $A$ and $B$ interact. Thus we say that $B$ interferes with $A$ in producing the event $Y_{=}1$ when $A$ and $B$ are both causal factors for $Y$, and there exists a possible intervention on $B$ which has the power of preventing any intervention on $A$ from causing the event $Y_{=}1$. This




*(a)* *(b)*




*(c)*

Figure 1: *(a)* our initial problem setting, *(b)* assumptions about the relationships between different regimes of data collection are added by the inclusion of *intervention indicators* in the graph, as discussed in Section 3, *(c)* the effects of $A$ and $B$ on $Y$ are jointly, but not individually, unconfounded.

can occur without also having $A$ interfering with $B$. We talk of *weak coaction* [resp., strong coaction] when at least one [resp., either] of $A$ and $B$ interferes with the other.

The above concepts are defined in terms of the behaviour of the system under a (real or hypothetical) *interventional* regime, where $A$ and $B$ are forced to take on specific value values. However in Section 6 we show that the proposed tests can be applied to data collected under under other regimes, *e.g.* observational. In Section 5, the conditions under which these tests are meaningful are studied in terms of conditional independence properties of an Augmented Directed Acyclic Graph (ADAG) representation of the problem ([5]). The ADAG will simultaneously represent the consensus causal theory about the system under study, and assumptions about the behaviour of the system across different regimes of data collection. ADAGs are briefly reviewed in Section 3. The scientific relevance of the method and its practicality in complex study designs are illustrated with the aid of two studies of the molecular determinants of coronary artery disease, one of numerous areas in biomedical research where an assumption of deep determinism could be defensible.

## 2 Interference and coaction

Henceforth we make the deep determinism assumption of Equation (1). The set of possible values of $A$ [resp., $B$, $V$] is denoted by $\mathcal{A}$ [resp., $\mathcal{B}, \mathcal{V}$].

**Definition 2.1 (Irrelevance)** *Factor $B$ is* (causally) irrelevant *to $Y$ in context $V{=}v$, given $A$, if $f(a,b,v) = f(a,b',v)$ for all $a \in \mathcal{A}, b, b' \in \mathcal{B}$.*

**Definition 2.2 (Interference)** *We say that $A$* interferes *with $B$ in producing the event $Y{=}1$ if, in some context $V{=}v$, $B$ is not irrelevant to $Y$ given $A$ and, for some $\hat{a} \in \mathcal{A}$ and all $b \in \mathcal{B}$,*

$$f(\hat{a}, b, v) = 0. \tag{2}$$

That is, in that context, there exists a value $\hat{a}$ such that, when we set $A{=}\hat{a}$, the event $Y{=}1$ will never happen, whatever value we impose on $B$.

**Definition 2.3 (Weak coaction)** *We say that $A$ and $B$* weakly coact *to produce the event $Y{=}1$ if at least one of $A$ and $B$ interferes with the other to produce the event $Y{=}1$.*

**Definition 2.4 (Strong coaction)** *We say that $A$ and $B$* strongly coact *to produce the event $Y = 1$ if each of $A$ and $B$ interferes with the other to produce the event $Y = 1$.*

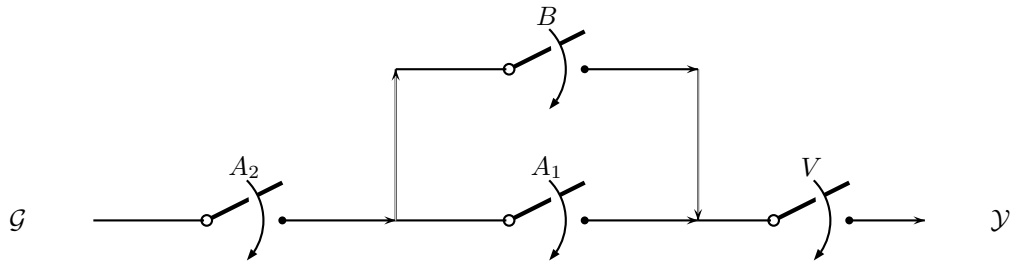



Figure 2: Electrical circuit illustration of coaction asymmetry. Imagine that an electrical voltage is applied between pins $\mathcal{G}$ and $\mathcal{Y}$. Let $Y = 1$ indicate absence of current between these two pins. Let $Y = 0$ indicate presence of current between these two pins. See main text for discussion of this example.

**Example (Logical):** Under a regime of intervention on variables $A \in \{0, 1, 2\}$ and $B \in \{0, 1\}$, let the binary response $Y$ depend on these two variables according to the logical law $Y = (A{=}2) \vee ((A{=}1) \wedge (B{=}1))$. Neither of $A$ or $B$ is irrelevant to $Y$. Setting $A$ to the value 0 will prevent the event $Y{=}1$, whatever the value we impose on $B$. However, when $A$ is set to value 2, event $Y{=}1$ will happen whatever value we impose on $B$. Hence $B$ does *not* interfere with $A$, while $A$ interferes with $B$, in producing the event $Y{=}1$. Thus, $A$ and $B$ coact weakly (but not strongly) in producing the event $Y{=}1$.

**Example (Electrical):** Consider the circuit of Figure 2, where we imagine an electrical voltage applied between pins $\mathcal{G}$ and $\mathcal{Y}$, and we take $Y = 1$ [resp., $Y = 0$] to indicate that current flows [resp., does not flow] between these two pins. Let the context variable be $U$, describing the unobserved state of the $U$–switch, each of the two possible states (OPEN, CLOSED) having positive probability. Let variable $A$ index the four possible configurations of the $A$–switches, and variable $B$ the position of the $B$–switch. The flow of current depends on the configuration of the switches via the well known deterministic laws of electrical circuits: this model thus satisfies deep determinism. Then in context $U{=}$ CLOSED, variable $B$ is not irrelevant to $Y$ since, when $A_1$ is open and $A_2$ closed, acting on $B$ will have an effect on current flow. However,

when $A_{2=}$ is open, no intervention on the $B$–switch can restore the current flow. Hence, in context $U_=$ CLOSED, variable $A$ interferes with $B$ in producing current flow.

**Example (Binary):** If $A$ and $B$ are binary, Equation (1) implies that, for a given value $v$ of $V$, the function $f$ takes one of sixteen possible patterns. First consider patterns $Y = TRUE$, $Y = FALSE$, $Y_{=}A$, $Y_{=}\overline{A}$, $Y_{=}B$ and $Y_{=}\overline{B}$. In all these patterns, at least one of $A$ or $B$ is irrelevant to $Y$, and therefore, by definition, neither of $A$ and $B$ interferes with the other in producing the event $Y_{=}1$. Next consider patterns $Y_{=}A \vee B$, $Y_{=}\overline{A} \vee B$, $Y_{=}A \vee \overline{B}$ and $Y_{=}\overline{A} \vee \overline{B}$, where the disjunctive form implies neither factor interferes with the other. Finally consider patterns $Y_{=}A \wedge B$, $Y_{=}\overline{A} \wedge B$, $Y_{=}A \wedge \overline{B}$, $Y_{=}\overline{A} \wedge \overline{B}$, $Y_{=}(A_{=}B)$ and $Y_{=}(A \neq B)$, where neither of $A$ and $B$ is irrelevant, and where no value of $A$ [resp., of $B$] produces the event $Y_{=}1$ unless $B$ [resp., $A$] takes on a particular value. Hence, in these last six patterns, each of $A$ and $B$ interferes with the other in producing the event $Y_{=}1$. We conclude that, in the special case where $A$ and $B$ are binary, there can be no interference asymmetry between $A$ and $B$: either they do or they do not interfere each with each other. Thus in this case weak and strong coaction coincide, and are essentially equivalent to the notion of interdependence given by [18].

**Example (Biological determinism):** Suppose a genetic mutation $A$ can induce a structural change in protein $\alpha$, causing disease $Y$ in certain individuals when the protein is expressed normally. Hence $A$ is not irrelevant to $Y$. Mutation $B$, located in the promoter region of the coding gene of $\alpha$, reduces the level of expression of $\alpha$. As a consequence, in the above individuals, presence of $B$ prevents any structural disfunctionality in protein $\alpha$ from causing the disease. In this case $B$ interferes with $A$ in causing disease $Y$ — an example of what geneticists call "epistasis".

We conclude this section with a remark. We have discussed "coaction to produce". We could similarly have defined "coaction to prevent". Coaction to prevent does not imply coaction to produce, nor *vice versa*. The scientific application and question of interest will usually dictate interest in one of the two directions.

# 3 Monotonicity

Sometimes we may be able to make assumptions about the ordering of the values of $Y$ in response to configurations of $A$ and $B$. In the electrical example of the previous section, for example, increasing the number of switches in CLOSED position can never cause the current flow to be switched off. Sometimes assumptions of this kind can be formulated as properties of *monotonicity*, as follows.

**Definition 3.1** *The effect of $A$ upon $Y$ is said to be* non-decreasing (with re-

spect to $B$) *if, for any configuration* $(b, v)$ *of* $(B, V)$*, the following implication holds:* $f(a, b, v) = 1$ *AND* $a' \geq a \Rightarrow f(a', b, v) = 1$.

**Definition 3.2** *The effect of* $A$ *upon* $Y$ *is said to be* non-increasing (with respect to $B$) *if, for any configuration* $(b, v)$ *of* $(B, V)$*, the following implication holds:* $f(a, b, v) = 0$ *AND* $a' \geq a \Rightarrow f(a', b, v) = 0$.

**Definition 3.3** *The effect of* $A$ *upon* $Y$ *is said to be* monotonic (with respect to $B$) *if it is either non-decreasing or non-increasing with respect to* $B$.

**Definition 3.4** *The effect of* $A$ *upon* $Y$ *is said to be* consistent (with respect to $B$) *if whenever, for any* $(a_1, a_2)$ *pair, the inequality* $f(a_1, b, v) \geq f(a_2, b, v)$ *holds for some* $(b, v)$ *configuration, it holds for all* $(b, v)$ *configurations.*

Clearly monotonicity implies consistency; conversely, under consistency we can re-order the values to yield monotonicity. [3] discuss the situation where a change in the value of $A$ may give rise to a reversal of the effect of $B$ upon outcome. Such *qualitative interaction* violates consistency. Some authors consider qualitative interaction to be interpretable in terms of mechanism. A formal test, different from the standard statistical test for departures from additivity, should be performed to assess whether a qualitative interaction could be due to chance variation. One such test has been proposed by [2]. The tests proposed in this paper, which also differ from standard statistical interaction tests, establish conditions for an interpretation of interaction in terms of mechanism without necessarily requiring that the underlying interaction be qualitative.

# 4 Augmented Directed Acyclic Graphs

Coaction has been defined under a (real or hypothetical) interventional regime. The tests for coaction we shall later propose may be applied more generally, such as when the data are observational. This, however, will require stringent assumptions, for example that $V$ be conditionally independent of $A$ and $B$ and of the way these two variables have been generated. In many applications it will be possible, and is then helpful, to represent such assumptions, in combination with further assumptions based on our causal understanding of the problem, by means of an Augmented Directed Acyclic Graph (ADAG).

Examples of ADAGs are given in Figure 1. Figure 1*(b)* is an ADAG specialisation of the simple problem setting of Figure 1*(a)*. An important feature of ADAGs is inclusion of *intervention indicators*, exemplified in Figures 1*(b)–(c)* by nodes $\sigma_A$ and $\sigma_B$. These nodes take values indicating the particular regime, observational or experimental, under which the values of a corresponding domain variable arise. With $A$ and $B$ binary, for example, each of $\sigma_A$ and $\sigma_B$ will have possible values in $(\emptyset, 0, 1)$, the interpretation being that, when $\sigma_A = \emptyset$, the variable $A$ is generated randomly by Nature, under the circumstances governing the observational data; while $\sigma_A = a \in \{0, 1\}$ indicates an interventional

setting in which value $a$ is imposed on $A$; and similarly for $B$. Although regime indicators are not random variables, we can still query the ADAG, using the *d-separation* criterion of [8], or the equivalent *moralisation* criterion of [10], to read off conditional independencies implied by the graph. These independencies will generally reflect properties of the system under study *and* judgements about the way we expect the system to behave under data collection regimes different from the actual one. The graphs of Figures 1*(b)* and 1*(c)*, for example, embody the conditional independence property, expressed in the notation of [7]: $Y \perp\!\!\!\perp (\sigma_A, \sigma_B) \mid (A, B)$, read as "$Y$ is conditionally independent of $(\sigma_A, \sigma_B)$, given $(A, B)$". This represents an assumed property of invariance across regimes: that once we know the values of $A$ and $B$, the distribution of $Y$ will not further depend on the regime of data collection, as represented by $(\sigma_A, \sigma_B)$. In other words, in these two examples, the distribution of $Y$ does not depend on the way the $(A, B)$ configuration has arisen, be it observationally or interventionally.

## 5 The core conditions

Identifiability conditions for mechanistic interaction are typically succinctly stated in terms of the effects of $A$ and $B$ on $Y$ having to be "unconfounded", conditional on some observed variable $C$. We adopt a different approach, assuming a consensus ADAG representation of the problem is available. Conditions for validity of the test proposed in the next section are then phrased in terms of conditional independence properties of the ADAG. This discipline allows us to be more precise in our claims than a formulation in terms of "no confounding". Another advantage of the ADAG-based approach is that it makes it easier to relate the conditions for applicability of a test to the substantive assumptions about the problem.

The assertion "the (joint) effects of $A$ and $B$ on $Y$ are unconfounded" might be interpreted as saying that there exists an observed variable $C$ such that the following two conditions are satisfied:

$$C \perp\!\!\!\perp \sigma \qquad \text{and} \qquad Y \perp\!\!\!\perp \sigma \mid (A, B, C) \tag{3}$$

where $\sigma := (\sigma_A, \sigma_B)$, with possible values $\sigma = (a, b)$, corresponding to setting $A = a, B = b$, and $\sigma = (\emptyset, \emptyset)$, also denoted by $\sigma = \emptyset$, when both $A$ and $B$ arise naturally. In this case we say that $C$ is a *sufficient covariate* for the joint effects of $A$ and $B$ on $Y$ ([9]). In accordance with the "back–door criterion" of [11], under these conditions the joint causal effect of $(A, B)$ on $Y$ will be estimable from observational data when $C$ is also observed. Note that these conditions need not imply that $C$ is sufficient for the individual causal effects of each of $A$ and $B$ on $Y$ (which would involve extending (3) to apply also when only one factor is intervened on, *i.e.* for $\sigma$ of the form $(a, \emptyset)$ or $(\emptyset, b)$). Thus in cases *(b)* and *(c)* of Figure 1, $C_{=}\emptyset$ is sufficient for the joint effects of $A$ and $B$ on $Y$, but is sufficient for the individual effects only in Figure 1 *(b)*.

However, neither sufficiency for the joint effects nor sufficiency for the individual effects is what we need to ensure applicability of the test of the next section for a general regime of data collection. In our analysis, the additional observable variable $C$ must be a function of the overall context variable $V$ featuring in the "deep determinism" property (1). Thus we can consider $V = (C, U)$, with $C$ observed and $U$ unobserved.

We shall require the simultaneous validity of the following four *core conditions*:

**Definition 5.1 (Core conditions)** *There exists a (possibly empty) set $C$ of observable context variables and a set $U$ of (typically unobserved) context variables such that:*

1. *(deep determinism) $Y_{=}f(A, B, C, U)$ for some deterministic function $f$, which is the same no matter how the variables $(A, B, C, U)$ are generated.*
2. $Y \perp\!\!\!\perp \sigma \mid (A, B, C, U)$
3. $U \perp\!\!\!\perp (A, B, \sigma) \mid C$,
4. $A \perp\!\!\!\perp B \mid (C, \sigma)$.

Whenever Condition 1 is satisfied, we say that $Y$ is *functional* with respect to $(A, B, C, U)$. Condition 2 essentially repeats the second part of Condition 1, but it is helpful to display it explicitly. Condition 3 says that, conditionally on $C$, variable $U$ has the same distribution in all regimes, and is independent of $A$ and $B$ (this will hold, in particular, if the full context variable $(C, U)$ has the same distribution in all regimes and is independent of $A$ and $B$); while Condition 4 requires $A$ and $B$ to be independent, given $C$, in the observational regime (this property necessarily holding when $A$ and $B$ are set by intervention).

The following theorem can be proved straightforwardly using general properties of conditional independence ([7], [11]).

**Theorem 5.1** *Core conditions 2 and 3 imply $Y \perp\!\!\!\perp \sigma \mid (A, B, C)$.*

Our core conditions imply the second condition of Equation (3), but not the first. It seems useful and instructive to discuss the differences between the two sets of conditions with the aid of examples. In the following examples interest focuses on testing coaction of variables $A$ and $B$ in producing the event $Y_{=}1$, based on observational data about variables $(A, B, Y)$ and, sometimes, a further variable $Z$.

Figures 1*(b)–(c)* satisfy the conditions of Equation (3) when $C_{=}\emptyset$. In both these examples, the distribution of $Y$ given $(A, B)$ does not depend on the way the configuration of values of $(A, B)$ is generated, be it observationally or by intervention. However, while Figure 1*(b)* satisfies the core conditions once we assume $Y$ to be functional with respect to $(A, B, U)$), Figure 1*(c)* violates core condition 4.

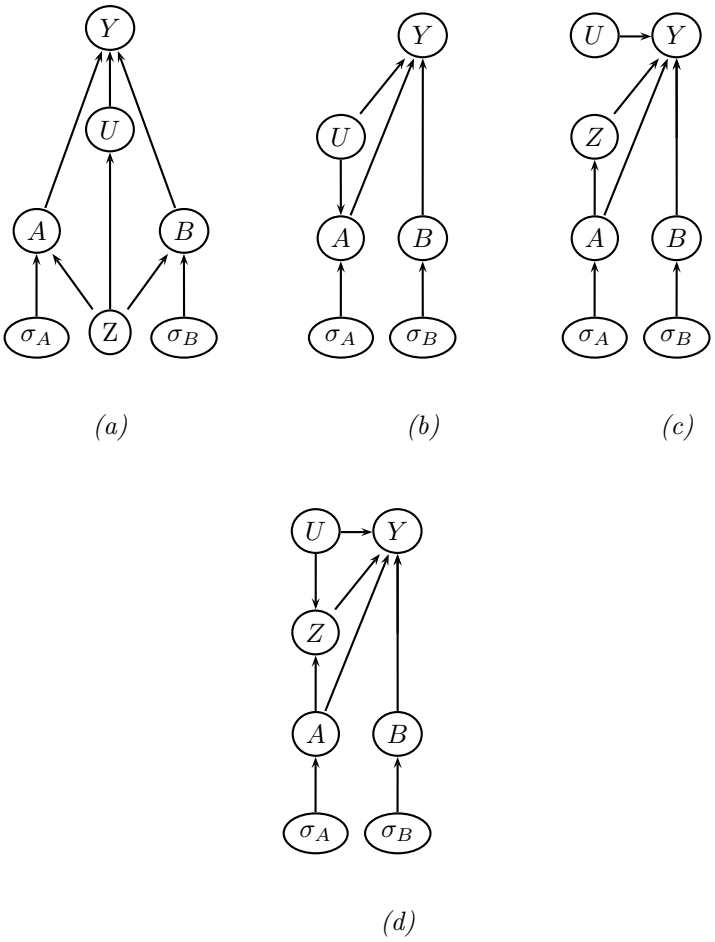



Figure 3: ADAG representations of problem examples discussed in the main text.

Now consider the example of Figure 3*(c)*. If the researcher engaged in a test of coaction between $A$ and $B$ follows the "no confounding" conditions of Equation (3), he or she will notice that these conditions are satisfied for $C_{=}\emptyset$, and might therefore proceed to perform the test without conditioning on $Z$. By contrast, if the researcher follows the core conditions of Definition 5.1, he/she will notice that ignoring $Z$ (that is, setting $C_{=}\emptyset$), is valid *only* under the assumption that $Y$ is functional with respect to $(A, B, U)$. This appears to be a tremendously stringent assumption, which we may accept only if, for every value of $U$, variable $Y$ is a deterministic function of $(A, B, Z)$ *and* $Z$ is a deterministic function of $A$. A more appropriate choice, according to the conditions of Definition 5.1, is to set $C_{=}Z$. The latter choice would make more sense from a further point of view, that is, it would test coaction of the effect of $B$ (on $Y$) and the *direct* effect of $A$ (on $Y$), unmediated by $Z$. In summary, in this example, the two sets of conditions lead to different choices, in the sense that the best choice according to the core conditions violates the "no confounding" conditions of Equation (3).

Many of the above considerations also apply to the example of Figure 3*(d)*. In particular, in this last example, setting $C_{=}\emptyset$ would appear a safe option according to the 'no confounding" conditions of Equation (3). And it would, in addition, satisfy core conditions *(2)* to *(4)*. A possible difficulty with this choice would however arise when negotiating core condition 1. In the light of core condition 1, choice $C_{=}\emptyset$ means we are ready to assume $Y$ to be deterministic when we condition on $(A, B, U)$, *but not* on $Z$. This is sensible only if we believe $Z$ to be *itself* is a deterministic function of its predecessors in the graph. Neither does the option $C_{=}Z$, in this example, solve the problem. For conditioning on $Z$ will typically introduce dependence between $U$ and $A$, violating core condition 3.

# 6 Testing coaction

We now present a test for coaction of variables $A$ and $B$ in producing the event $Y_{=}1$, assuming that there exists a (possibly empty) set $C$ of observed variables such that the core conditions of the previous section are valid. We allow $A$ and $B$ to be ordered categorical or continuous variables, in any combination. If either variable is not binary, we consider some dichotimisation of its range. Thus for $A$ we would choose a threshold $\tau_A$ and define $\alpha := \{a \in \mathcal{A} : a > \tau_A\}$, $\overline{\alpha} := \{a \in \mathcal{A} : a \leq \tau_A\}$. Similarly for $B$ we would have $\tau_B, \beta, \overline{\beta}$. We also use $\alpha$ to denote the truth-value (0 or 1) of the event $A \in \alpha$, *etc*.

In the sequel, all probabilities are computed under the observational regime $\sigma = \emptyset$.

For $i, j = 0, 1$, let

$$R_{ijc} := P(Y{=}1 \mid \alpha{=}i, \beta{=}j, C{=}c)$$
$$\overset{\text{core condition 3}}{=} \int_u R_{ijc}(u)\; P(u \mid C{=}c)\; \mathrm{u}$$

where, for any value $u$ of $U$,

$$\begin{aligned}
R_{ijc}(u) &:= P(Y{=}1 \mid \alpha = i, \beta = j, C{=}c, U{=}u) \\
&= \int_{\alpha=i} \mathrm{a} \int_{\beta=j} \mathrm{b}\; P(Y{=}1 \mid a, b, C{=}c, U{=}u)\; P(a, b \mid \alpha = i, \beta = j, C{=}c, U{=}u) \\
&\overset{\text{core conditions 1, 2}}{=} \int_{\alpha=i} \mathrm{a} \int_{\beta=j} \mathrm{b}\; f(a, b, c, u)\; P(a, b \mid \alpha = i, \beta = j, C{=}c, U{=}u) \\
&\overset{\text{core condition 3}}{=} \int_{\alpha=i} \mathrm{a} \int_{\beta=j} \mathrm{b}\; \frac{f(a, b, c, u)\; P(a, b \mid C{=}c)}{P(\alpha = i, \beta = j \mid C{=}c)} \\
&\overset{\text{core condition 4}}{=} \int_{\alpha=i} \mathrm{a}\, P(a \mid \alpha = i, C{=}c) \int_{\beta=j} \mathrm{b}\; f(a, b, c, u)\; P(b \mid \beta = j, C{=}c).
\end{aligned} \tag{4}$$

**Definition 6.1** *Variable $A$ is said to be* $\alpha$-insensitive *with respect to $Y$ if the following implication is valid for all $(b, c, u)$:*

$$\textit{IF}\;\; f(a, b, c, u) = 0 \quad \textit{for some}\; a \in \alpha\; \textit{AND}\; a' \geq a \quad \textit{THEN} \quad f(a', b, c, u) = 0 \tag{5}$$

We similarly define the $\beta$-insensitivity property for $B$. Trivially $\alpha$-insensitivity holds if $\alpha$ consists of a single point. We are now ready to state the main theorem:

**Theorem 6.1** *Let the binary outcome variable $Y$ depend on observed variables $(A, B, C)$ and on unobserved variable $U$, where $A$ and $B$ are allowed to be ordered categorical or continuous, in any combination of these two types. Let the effect of $A$ [resp., $B$] upon $Y$ be monotonic with respect to $B$ [resp., $A$], and suppose that, for some dichotomizations of $A$ and $B$, and some value $c$ of $C$:*

$$R_{11c} - R_{10c} - R_{01c} > 0. \tag{6}$$

*Then under the core conditions and the $\alpha$–insensitivity property for $A$, variable $B$ interferes with $A$ in producing the event $Y = 1$. Similarly, whenever the $\beta$–insensitivity property holds for $B$, variable $A$ interferes with $B$ in producing the event $Y = 1$; in either case $A$ and $B$ weakly coact to produce the event $Y{=}1$.*

**Proof** Equation (6) can be expressed as

$$\int [R_{11c}(u) - R_{10c}(u) - R_{01c}(u)]\, p(\mathrm{u} \mid C{=}c) > 0. \tag{7}$$

It follows that there is a positive probability of obtaining a value $u^*$ of $U$ such that $R_{11c}(u^*) - R_{10c}(u^*) - R_{01c}(u^*) > 0$; in particular, $R_{11c}(u^*) - R_{10c}(u^*) > 0$. Thus, using (4),

$$\int_{a\in\alpha} \dot{a}\, P(a \mid A \in \alpha, C{=}c) \left[ \int_{b\in\beta} \dot{b}\, f(a,b,c,u^*)\, P(b \mid B \in \beta, C{=}c) \right.$$
$$\left. - \int_{b\in\overline{\beta}} \dot{b}\, f(a,b,c,u^*)\, P(b \mid B \in \overline{\beta}, C{=}c) \right] \qquad > 0,$$

from which it follows that there exists a value $a_1 \in \alpha$ such that

$$\int_{b\in\beta} \dot{b}\, f(a_1,b,c,u^*)\, P(b \mid B \in \beta, C{=}c) > \int_{b\in\overline{\beta}} \dot{b}\, f(a_1,b,c,u^*)\, P(b \mid B \in \overline{\beta}, C{=}c). \quad (8)$$

Since the left-hand-side of the above inequality is thus positive, and $f = 0$ or 1, we must have

$$f(a_1, b_1, u^*, c) = 1 \qquad \text{for some } \; b_1 \in \beta. \quad (9)$$

Also we cannot have have $f(a_1, b, u^*, c) = 1$ for all $b \in \overline{\beta}$, since in this case the right-hand side of (8) would equal 1, whereas the left-hand side can not exceed 1. We deduce that

$$f(a_1, b_2, u^*, c) = 0, \qquad \text{for some } \; b_2 \in \overline{\beta}. \quad (10)$$

Because Equation (7) is symmetrical in $A$ and $B$, we similarly obtain:

$$f(a_2, b_3, u^*, c) = 1, \qquad \text{for some } \; a_2 \in \alpha \; \text{ and } \; b_3 \in \beta, \quad (11)$$
$$f(a_3, b_3, u^*, c) = 0, \qquad \text{for some } \; a_3 \in \overline{\alpha}. \quad (12)$$

Under the assumed monotonicity of the effect of $B$ upon $Y$, and remembering that $\beta$ lies above $\overline{\beta}$, Equations (9)–(10) imply that $f$ is non-decreasing with $B$ for any configuration of $(A, C, U)$. Similarly, Equations (11)–(12) imply that $f$ is non-decreasing with $A$ for any configuration of $(A, C, U)$. Equations (9) (10) (11) and (12) tell us that there is a context $(U, C) = (u^*, c)$ where variables $A$ and $B$ are *not* irrelevant to $Y$ with respect to each other. Then according to Definition 2.2, in order to prove that $B$ interferes with $A$ in producing the event $Y{=}1$, we only need prove that, for some value imposed on $B$, no value of $A$ will produce the event $Y{=}1$, that is:

$$f(a, b_2, u^*, c) \; = \; 0 \quad \forall a. \quad (13)$$

In fact, the following two implications follow from Equation (10):

$$a^* < a_1 \;\Rightarrow\; f(a^*, b_2, u^*, c) \overset{f \text{ non-decreasing with } A}{=} 0$$
$$a^* \geq a_1 \;\text{ AND }\; a_1 \in \alpha \;\Rightarrow\; f(a^*, b_2, u^*, c) \overset{A \text{ is } \alpha\text{-insensitive wrt } Y}{=} 0$$

from which Equation (13) follows. We then conclude that, under an assumed $\alpha$–*insensitivity condition* for $A$, Equation (6) implies that variable $B$ interferes with $A$ in producing the event $Y = 1$. Similarly we can prove that, under an assumed $\beta$–insensitivity property for $B$, variable $A$ interferes with $B$ in producing the event $Y = 1$, which completes the proof.

**Remark 1:** The theorem holds also in those situations where we can have its conditions satisfied by an appropriate recoding of $A$ and $B$.

**Remark 2:** The theorem can be applied conditional on the generic individual belonging to a particular population stratum defined on the basis of $(A, B)$.

The following example illustrates the two remarks above. Consider a discrete variable $A \in \{1, 2, 3, 4\}$ and a continuous variable $B$ on the $(0, 1)$ interval. Assume monotonicity of the effects of $A$ and $B$ on $Y$, and let us restrict attention to the stratum of individuals with $A > 1$. Let us then recode variable $A$ by setting $A^* := 5 - A$. Then suppose for $\alpha : \{A^*{=}3\}$ and $\beta : \{B > 0.5\}$ that the data strongly support the inequality $R_{11} - R_{01} - R_{10} > 0$. Because $\alpha$ consists of a single point, and therefore $A$ is $\alpha$-insensitive with respect to $Y$, we may conclude that $B$ interferes with $A$ in producing the event $Y{=}1$. The reverse inference, that $A$ interferes with $B$, is possible if $B$ is $\beta$-insensitive with respect to $Y$, but this assumption may be problematic since $\beta$ does not consist of a single point.

# 7 Examples

We discuss the examples of Figures 3*(a)—(b)*.
**Example of Figure 3*(a)*** Let $Y$ be an indicator of disease, depending on a pair $(A, B)$ of genetic variants in linkage equilibrium with each other; and let covariate $Z$, representing genealogical information, say, be sufficient for the effects of $A$ and $B$ on $Y$. Then the graph of Figure 3*(a)* might be an acceptable representation of the problem. Suppose further there is consensus that $Y$ is functional with respect to $(A, B, U)$, for example because the effects of the two variants on $Y$ are thought to operate through a common molecular mechanism. Then the core conditions are satisfied if we take $C \equiv Z$, and so observational $(A, B, Z, Y)$ data can be used to test for $A$–$B$ coaction to produce the event $Y{=}1$.
**Example of Figure 3*(b)*** In this example, where $C$ is necessarily empty, node $U$ is *not* independent of $A$, which violates core condition 3. Consequently a set of observational $(A, B, Y)$ data will typically not suffice for us to be able to test productive coaction of $A$ and $B$ by using the proposed method.

# 8 Relations with previous work

In certain formal frameworks for "statistical causality", including Pearl's structural equation formulation ([11], chapter 7) and the potential response framework of [14], it is possible to construct a totally fictitious mathematical variable $V$ which makes (1) true by mathematical fiat. Our approach differs in that we conceive of the context variable $V$ as both real and relevant — and thus in principle observable; its relationships with the remaining variables in the problem need be negotiated and explicitly represented in the causal model. This has

practical consequences for data analysis. Consider, for example, Figure 3*(c)* and *(d)*. These two examples differ only in that, in the former, on the basis of contextual knowledge, we judge the unobserved "context variables" $U$ that differentiate the possible behaviours of $Y$ in response to $(A, B)$ to be *a priori* unrelated to $Z$, whereas in the latter example, these unknown variables are judged to act as unobserved confounders of the effect of $Z$ upon $Y$. We have seen that this difference has consequences on our decision to apply the method, and whether or not we should condition on $Z$.

Also, our method replaces the generic assumption of "the effect of $A$ and $B$ on $Y$ is not confounded given $C$" with a formal set of independencies (the core conditions) that need to be satisfied by the causal model. We have seen in Section 5 that this formal method can capture important differences between different applications.

# 9 Illustrative study: rs1333040 coacts with statins

Within the Italian genetic study of early-onset myocardial infarction ([1]), between 1996 and 2002, an incident sample of 2050 cases was selected on the basis of an hospitalization for myocardial infarction (MI) between age 40 and age 45, over a set of 125 Coronary Care Units spread nationwide. After entering the study, each sample subject produced a blood sample from which plasma was separated and DNA extracted, and was then prospectively monitored for an average of 12 years of follow-up. Let the outcome of the follow-up be represented by a binary variable, $Y$, indicating whether a re-infarction or cardiovascular death were observed ($Y$=1), or not observed ($Y$=0) within a period of 120 months from study entry.

The research group agrees on the assumptions represented in the ADAG of Figure 4. According to the graph, each case is characterized by the following variables. Variable $G$ is a function of the genotype at rs1333040, a single nucleotide polymorphism (SNP) located in chromosomal region 9p21.3. We define $G$ to take value 1 in presence of two copies of the major rs1333040 allele, and value 0 otherwise. Variable $Z$ is the severity of coronaropathy at study entry. Variable $T$ is the calendar year at study entry. Variable $U$ represents a set of unknown confounders. Variable $S$ indicates whether the subject was assigned to statin treatment right after study entry ($S$=1) or never after study entry ($S$=0), and $I$ indicates presence/absence of hypercholesterolemia at study entry. Variable $T$ here acts as a surrogate for relevant factors that vary with calendar time. These include therapy evolution, progress of medical knowledge and impact of legislation. These factors are assumed to influence both medical practice, specifically concerning use of statins, and the clinical outcome $Y$. During the study period, National Guidelines concerning use of statins had not yet come into force, and the decision whether or not to administer statins to patients of the kind we are studying was taken more or less randomly by the recruiting Coronary Care Unit, though to some extent dependent on whether or

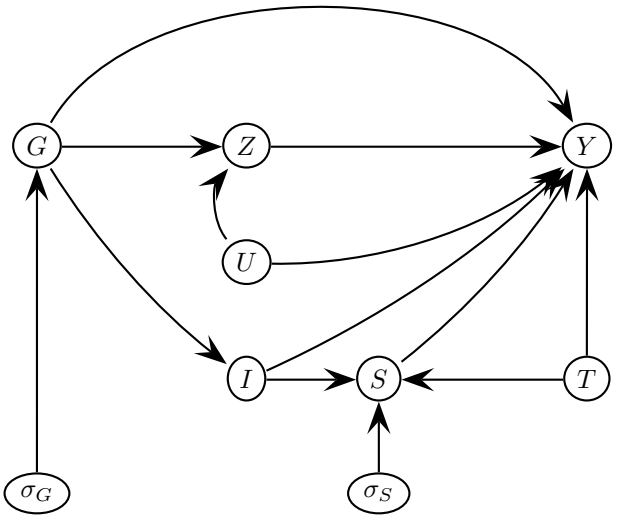



Figure 4: ADAG representation of our illustrative study of coaction between a gene tagged by single nucleotide polymorphism rs1333040 and statin treatment in producing myocardial infarction. See main text for a justification of the causal relationships depicted in this graph.

not the patient was found to have hypercholesterolemia at study entry. This is accounted for in the graph by the $I \to S$ arrow. The graph also conservatively allows that susceptibility to hypercholesterolemia may depend on the genotype at the SNP of interest, although evidence in support of this has never been found.

Instead of performing separate analyses within strata of $(T, I)$, we restrict analysis to the stratum of patients with hypercholesterolemia ($I_{=}1$), and assume that, in this stratum, the effect of $T$ does not interact with $G$ and $S$. We then model the effect of $(G, S, T)$ on $Y$ in the stratum of patients with $I = 1$ via the following linear risk Bernoulli model:

$$\left\{ \begin{array}{lcl} Y & \sim & \text{Bernoulli}(\pi), \\ \pi & = & \alpha + \phi_{S=0} + \phi_{G=1} + \gamma_{(S=0)\times(G=1)} + \delta_t \, T, \end{array} \right. \tag{14}$$

where $\delta_t$ represents a linear effect of calendar year, in years since 1970. If our data provide evidence of a departure of parameter $\gamma_{(S=0)\times(G=1)}$ from zero,

we say that variables $S$ and $G$ interact *statistically* in the stratum of hypercholesterolemic patients and, indeed, the results shown in Table 1 support this conclusion. The data seem to tell us that the beneficial effect of statins, in terms of reduction of risk of re-infarction in a hypercholesterolemic patient, is stronger in patients with $G = 0$. And that the highest risk is found in those hypercholesterolemic patients with $G = 0$ who do not receive statins.

Let's now shift from predictive to mechanistic inference, by examining whether the interaction between variables $S$ and $G$ can be upgraded from "statistical" to "mechanistic". In order to do this, we need to use a different statistical test, and to be explicit about the set of (fairly strong) assumptions discussed in the previous section. One of these is monotonicity of the effects, which appears to be reasonable, since it does not require prior knowledge of the "deleterious" allele of the SNP. Next, we need to assume that the core conditions hold. Define $C = (T, I)$. With this choice, core conditions 2 to 4 are satisfied, although core condition 1 — that $Y$ be a deterministic function of $(G, S, T, I, U)$ — could be problematic here unless we assume that, for any given value of $U$, variable $G$ influences $Z$ and $Y$ through the same molecular mechanism whereby interference with the effect of statin takes place. After accepting the core conditions, in accordance with the theory of Section 6, we partition the possible values of the rs1333040 genotype into the set $\alpha$ and its complement $\overline{\alpha}$. We define $\alpha$ to indicate presence of two copies of the most frequent rs1333040 allele, corresponding to $G = 1$, so that $\overline{\alpha}$ will represent the remaining two genotypic categories. We define $\beta$ to indicate that the patient is given statins, corresponding to $S = 1$, and we define $\overline{\beta}$ to indicate that the patient is *not* given statins, corresponding to $S = 0$. Since each of $\alpha$ and $\beta$ contains just one value of the corresponding variable, $\alpha$-insensitivity and $\beta$-insensitivity hold in this case.

It is easy to show that, given $T = t$, the above model implies $R_{11t} - R_{10t} - R_{01t} = \gamma_{(S=0)\times(G=1)} - \alpha - \delta_t t$. This quantity, according to Table 1, is significantly greater than zero for all relevant values of $T$. Hence, in the light of our theory and under the assumptions discussed above, we conclude that $G$ and $S$ strongly coact to produce re-infarction. The interpretation may be phrased in a number of ways. One is to say that there exists some context in which hypercholesterolemic patients with the $G = 1$ genotype are safe from re-infarction, whether or not they take statins, whereas those with $G = 0$ develop or avoid re-infarction depending on assumption of statins. A counterfactual rephrasing of this is to say that some patients with $G = 0$, who developed re-infarction, would not have developed it, had they received statins. All this can be interpreted to suggest that statins and some gene tagged by rs1333040 influence susceptibility to reinfarction through a common pathway, which motivates a future effort to identify which gene is this, and what is its function. Some researchers might have got to the same conclusions from the results of the regression analysis, without consideration of the theoretical framework proposed in this paper. In our opinion, that would be careless. Not only do such conclusions require a statistical test of the kind proposed in this paper, which differs from a standard interaction test,

but also, they require explicit consideration of the (fairly strong) assumptions we have discussed in this paper.

## 10 Illustrative study: rs4620585 coacts with smoking

Each of the cases in the study of the previous section was paired with a control, matched by age and geographical region of origin. After eliminating individuals with missing data, 1666 controls remained available for the analysis. In this section we concentrate on "smoking habit", a binary indicator obtained by dichotomizing an (imprecisely recorded) daily number of cigarettes. We tested for possible coaction between smoking habit and one or more SNPs of a list of ten candidates from an independent study, an interesting signal being found at SNP rs4620585 of human chromosome 1, never previously been associated with a disease. The remaining discussion restricts attention to SNP rs4620585. Let $A$ signify rare rs4620585 homozygosity (RRH), and $B$ signify "smoker". Let $Y$ represent occurrence of early MI. We assume that the core conditions, and in particular condition 4, hold in this problem, once we assume (in accord with current knowledge) that the gene implicated by rs4620585 has no influence on smoking habit or addiction to nicotine.

On the basis of our data, we performed a linear-odds regression of the case-control indicator on SNP rs4620585 and smoking habit. This analysis yielded the estimated coefficients of Table 2. Because our "early MI" endpoint is rare, we may safely assume that the selection effect implicit in the case-control study affects the interaction parameter $\gamma$ and the intercept $\alpha$, in principle estimable only through a prospective study, only through multiplication by a common, unknown, positive constant. Hence we may take positivity of $(\gamma - \alpha)$ to imply positivity of the linear combination $R_{11} - R_{01} - R_{10}$ of prospective risks. Since Table 2 shows the quantity $R_{11} - R_{01} - R_{10}$ to be significantly greater than zero (no multiple testing adjustment), we conclude in favour of a potential mechanistic interaction between SNP rs4620585 and smoking. One interpretation of this result is to say that there are circumstances in which some patients, by virtue of a beneficial variant tagged by rs4620585, are safe from an early MI regardless of their smoking, whereas patients without that variant, who in the same circumstances developed an early MI, would have avoided it, had they *not* smoked.

## 11 Discussion

Statistical interaction — departure from some parametric model of independent effects of explanatory variables — is not necessarily interpretable as reflecting an underlying mechanism, not least because most statistical models are mathematical fictions ([4]). This is especially true when the modeller has to negotiate

continuous explanatory variables. Our proposed sufficient conditions for declaring coaction between continuous variables do not invoke specific parametric forms of dependence, and appear to provide a better basis for inference about mechanistic interaction. The proposed method does however, rely on the assumption that the mechanism studied is, at some deep level, deterministic — which is by no means universally appropriate, as shown by [6]. This assumption can, however, be defensible in some fields of application, and our choice of an illustrative study in molecular medicine reflects such concerns.

Finally, we would re-iterate that, unlike previous approaches to the problem, the proposed method avoids artificial mathematical constructs based on a potential response paradigm of statistical causality. While some of our tests are mathematically similar to previously proposed tests based on "principal stratum" arguments, our tests differ in that we insist the context variable $V$ be both real and relevant. Although $V$ may be wholly or partly unobserved, it is important in our method that it be, in principle at least, observable, and that its relationships with the remaining variables in the problem explicitly represented in the causal model. With the aid of study examples, we have shown that such an exercise is necessary to differentiate situations in which the method is applicable from situations in which it is not.

# References


[1] Diego Ardissino and Carlo *et al* Berzuini. Influence of 9p21.3 genetic variants on clinical outcomes in early-onset myocardial infarction. *Accepted for publication on the Journal of the American Cardiology Association*, 2010.

[2] A. Azzalini and D. R. Cox. Two new tests associated with analysis of variance. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(2):335–343, 1984.

[3] Amy Berrington de Gonzalez and David R Cox. Interpretation of interaction: a review. *The Annals of Applied Statistics*, 1(2):371385, 2007.

[4] David G. Clayton. Prediction and interaction in complex disease genetics: Experience in type 1 diabetes. *PLoS Genet*, 5(7):e1000540, 07 2009.

[5] A. P. Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70(2):161–189, 2002.

[6] A. P. Dawid. Probability, causality and the empirical world: A bayes-de finetti-popper-borel synthesis. *Statistical Science*, 19(1):44–57, 2004.

[7] A.P. Dawid. Conditional independence in statistical theory. *J. R. Stat. Soc., Ser. B*, 41:1–31, 1979.

[8] Dan Geiger, T. Verma, and Judea Pearl. Identifying independence in bayesian networks. *Networks*, 20(5):507–534, 1990.

[9] Hui Guo and Alexander Philip Dawid. Sufficient covariates and linear propensity analysis. In Yee Whye Teh and D. Michael Titterington, editors, *Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics, (AISTATS) 2010, Chia Laguna, Sardinia, Italy, May 13-15, 2010. Journal of*

*Machine Learning Research Workshop and Conference Proceedings*, volume 9, pages 281–288, 2010.

[10] Steffen Lilholt Lauritzen, A. Philip Dawid, B.N Larsen, and H.G. Leimer. Independence properties of directed markov fields. *Networks*, 20(5):491–505, 1990.

[11] Judea Pearl. *Causality : Models, Reasoning, and Inference*. Cambridge University Press, March 2000.

[12] K. J. Rothman. Causes. *American Journal of Epidemiology*, 104:587–592, 1976.

[13] K. J. Rothman and Sanders Greenland. Modern epidemiology. 1998.

[14] Donald Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100:322–331, 2005.

[15] Anders Skrondal. Interaction as departure from additivity in case-control studies: A cautionary note. *American Journal of Epidemiology*, 158(3), 2003.

[16] Tyler J. VanderWeele. Sufficient cause interactions and statistical interactions. *Epidemiology*, 20:6–13, 2009.

[17] Tyler J. VanderWeele. Sufficient cause interactions for categorical and ordinal exposures with three levels. *Biometrika*, 97(3):647–659, 2010.

[18] Tyler J. VanderWeele and James M. Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):4961, 2008.

[19] Tyler J. VanderWeele and James M. Robins. Minimal sufficient causation and directed acyclic graphs. *Ann. Stat.*, 37(3):1437–1465, 2009.

| | Estimate | Std. Error | z value | $p$-value |
|---|---|---|---|---|
| $\alpha$ (intercept) | -2.33 | 0.5 | -4.66 | $3e^{-6}$ |
| $\phi_{G=1}$ (wild-type rs1333040 homozygous) | -0.06 | 0.19 | -0.31 | 0.7 |
| $\phi_{S=0}$ (no statin treatment) | 1.41 | 0.24 | 5.85 | 4e-09 |
| $\delta_T$ (linear effect of calendar year - 1970) | -0.02 | 0.017 | -1.52 | 0.12 |
| $\gamma_{(S=0)\times(G=1)}$ | -1.0 | 0.33 | -3.0 | 0.002 |

Table 1: Parameter estimates from a linear-odds regression of the prospective binary endpoint in our illustrative study (re-infarction within six years from the index infarction) upon variables $S$ (the statin treatment indicator) and $G$ (a function of the genotype at SNP rs1333040). Variable $G$ is coded to take value 1 if the individual carries two copies of the most frequent allele at single nucleotide polymorphism rs1333040. This table reports estimates for the parameters of the regression model, as obtained from an analysis of 1200 subjects who were hospitalized on the basis of a myocardial infarction between 40 and 45 years of age, and were found at that point to have hypercholesterolemia. These estimates suggest that, in patients with hypercholesterolemia, statins decrease the risk of re-infarction regardless of the rs1333040 genotype ($G$), although their effect is stronger in patients with $G = 0$. At highest risk are those hypercholesterolemic patients with $G = 0$ who do not receive statins. Because the quantity $\gamma - \alpha$ is significantly greater than zero, we deduce that $G$ and $S$ interfere with each other (and hence strongly coact) to produce re-infarction.

| | Estimate | Standard Error | z value | $p$-value |
|---|---|---|---|---|
| Intercept ($\alpha$) | 0.25 | 0.013 | 18.7 | $< 2e^{-16}$ |
| smoker | 1.46 | 0.044 | 33.1 | $< 2e^{-16}$ |
| rare rs4620585 homozygous (RRH) | 0.07 | 0.056 | 1.2 | 0.19 |
| smoker $\times$ RRH ($\gamma$) | 0.9 | 0.22 | 4.09 | $4e^{-5}$ |

Table 2: Parameter estimates from a linear-odds regression of the early MI indicator upon "smoking habit", obtained by dichotomizing an original "Daily number of cigarettes" variable and genotype at SNP rs4620585, based on the set of cases and controls of our Illustrative study.