Published in final edited form as:

J R Stat Soc Series B Stat Methodol. 2012 March ; 74(2): 223–244. doi:10.1111/j.
1467-9868.2011.01011.x.

Semiparametric tests for sufficient cause interaction

Stijn Vansteelandt,

Department of Applied Mathematics and Computer Sciences Ghent University, 281 (S9)
Krijgslaan, 9000 Ghent, Belgium

Tyler J. VanderWeele, and

Departments of Biostatistics and Epidemiology Harvard School of Public Health, Boston, MA
02115, U.S.A.

James M. Robins

Departments of Biostatistics and Epidemiology Harvard School of Public Health, Boston, MA
02115, U.S.A.

Abstract

A sufficient cause interaction between two exposures signals the presence of individuals for whom the outcome would occur only under certain values of the two exposures. When the outcome is dichotomous and all exposures are categorical, then under certain no confounding assumptions, empirical conditions for sufficient cause interactions can be constructed based on the sign of linear contrasts of conditional outcome probabilities between differently exposed subgroups, given confounders. It is argued that logistic regression models are unsatisfactory for evaluating such contrasts, and that Bernoulli regression models with linear link are prone to misspecification. We therefore develop semiparametric tests for sufficient cause interactions under models which postulate probability contrasts in terms of a finite-dimensional parameter, but which are otherwise unspecified. Estimation is often not feasible in these models because it would require nonparametric estimation of auxiliary conditional expectations given high-dimensional variables. We therefore develop ‘multiply robust tests’ under a union model that assumes at least one of several working submodels holds. In the special case of a randomized experiment or a family-based genetic study in which the joint exposure distribution is known by design or Mendelian inheritance, the procedure leads to asymptotically distribution-free tests of the null hypothesis of no sufficient cause interaction.

Keywords

Double robustness; Effect modification; Gene-environment interaction; Gene-gene interaction; Semiparametric inference; Sufficient cause; Synergism

1 Introduction

Developments in genetic epidemiology have recently stimulated a large interest in tests for interaction or effect modification (see e.g. Maity et al., 2009; Murcray, Lewinger and Gauderman, 2009; Tchetgen and Robins, 2009; VanderWeele, 2010b; Vansteelandt et al., 2008a). In this article, we focus more broadly on studies whose goal is to determine whether

two exposures interact in the effect that they produce on a dichotomous outcome, e.g. disease status. Such interactions are commonly assessed by specifying logistic regression models in which the linear predictor includes the product between the exposures and then testing whether the corresponding regression coefficient is zero. While frequently used, these model-based tests are not entirely satisfactory because they are not scale-invariant (Mantel et al., 1977; Greenland, 1993): two variables that have an interaction under one statistical model (e.g. with a logistic link), may not have an interaction under a different model (e.g. with a log link). In particular, the presence of an interaction between two factors in a regression model does not necessarily signal the presence of individuals who get the disease only under certain values of both exposures (Rothman, 1976; Cordell, 2002; VanderWeele, 2009; Greenland, 2009). Standard model-based tests for interaction therefore do not allow one to draw inference about mechanistic interaction processes.

When the outcome and exposures are dichotomous, there is a natural, scale-invariant way in which to assess the presence of interactions between two exposures, based on the sufficient cause framework (Rothman, 1976; Koopman, 1981; VanderWeele and Robins, 2007). Specifically, consider an outcome Y with two binary causes A_1 and A_2 . Let $Y_{a_1 a_2}(\omega)$ denote the counterfactual value of Y for individual ω if A_1 were set to a_1 and A_2 were set to a_2 . Suppose there were an individual ω for whom $Y_{11}(\omega) = 1$ but $Y_{10}(\omega) = Y_{01}(\omega) = 0$; it would then be the case that for this individual the outcome would occur if both exposures were present but not if only one of the two were present. The presence of such an individual would thus signal the existence of a mechanism for the outcome that would operate if both exposures were present but not if only one of the two were present. We then say that a sufficient cause interaction is present between the exposures A_1 and A_2 .

VanderWeele and Robins (2007, 2008) provided empirical tests for sufficient cause interactions. Let $\mathbf{X}$ be some set of variables so that conditioning on $\mathbf{X}$ suffices to control for confounding for the estimation of the causal effect of $\mathbf{A} = (A_1 A_2)'$ on Y (i.e. so that $Y_a \perp\!\!\!\perp \mathbf{A} / \mathbf{X}$ for all $a = (a_1 a_2)'$ in the sample space Ω of $\mathbf{A}$). Further, define $\pi(a, \mathbf{X}) = E(Y_a / \mathbf{X}) = E(Y / \mathbf{A} = a, \mathbf{X})$ for each $a \in \Omega$. VanderWeele and Robins (2007, 2008) then show that if the probability contrast

$$\pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) - \pi((1, 0), \mathbf{X}) > 0 \quad (1)$$

for some $\mathbf{X}$, it entails the presence of a sufficient cause interaction between A_1 and A_2 . They further show that, when the exposures A_1 and A_2 have positive monotonic effects on the outcome Y in the sense that $Y_{a_1 a_2}(\omega) \geq Y_{a'_1 a_2}(\omega)$ whenever $a_1 \geq a'_1$ and $Y_{a_1 a_2}(\omega) \geq Y_{a_1 a'_2}(\omega)$ whenever $a_2 \geq a'_2$, then if the probability contrast

$$\pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) - \pi((1, 0), \mathbf{X}) + \pi((0, 0), \mathbf{X}) > 0 \quad (2)$$

for some $\mathbf{X}$, it implies that a sufficient cause interaction is present. Under such monotonicity assumptions, testing for a sufficient cause interaction is thus tantamount to testing for effect modification on the risk difference scale. Note that the probability contrast in (2) exceeds that in (1), so tests of sufficient cause interaction will be more powerful when one is willing

to assume positive monotonic effects (see Section 1 of the Online Supplement for further details).

This work on sufficient cause interactions has been extended to an arbitrary number of binary variables (VanderWeele and Richardson, 2009), to ordinal and categorical exposures (VanderWeele, 2010c) as often considered in genetic applications, and to tests for "compositional epistasis" (Cordell, 2002; Phillips, 2008; VanderWeele 2010ab) in which the effect of one genetic factor is masked by another. In all the above work, a test for the presence of sufficient cause interaction is tantamount to a test of the composite null hypothesis that, for a particular known function $\lambda(\mathbf{a})$,

$$H_0: \sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}) \leq 0 \quad \text{w.p.1} \quad (3)$$

versus the alternative hypothesis

$$H_1: \text{the event } \sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}) > 0 \text{ has positive probability,} \quad (4)$$

where $\mathbf{A}$ and $\mathbf{a}$ are now vectors with possibly more than two components, each of which takes values in a finite set. In the case (1), $\lambda(1, 1) = 1$, $\lambda(1, 0) = \lambda(0, 1) = -1$, $\lambda(0, 0) = 0$. The case (2) differs only in that $\lambda(0, 0) = 1$.

The problem of testing the above hypotheses H_0 versus H_1 is important in testing for sufficient cause interactions and constitutes a statistical problem of interest in its own right. In principle, one could carry out such testing by modelling $\pi(\mathbf{A}, \mathbf{X})$ via Bernoulli regression models with linear link (Greenland, 2004), but these may easily be misspecified as they do not ensure that the expected dichotomous outcomes stay between 0 and 1. Likewise, testing the null hypothesis (3) on the basis of contrasts between predicted values obtained from logistic regression models is not entirely satisfactory, because such models are highly non-additive in typical settings. By thus a priori imposing non-additivity under parsimonious, but misspecified logistic regression models, one would therefore induce a bias towards non-additivity on the risk difference scale, leading to an inflation of the Type I error (Greenland, 2009). In addition, a drawback of logistic regression modeling is that it requires separate tests at each $\mathbf{X}$ -level (Skrodingal, 2003) because the contrast

$$\sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}) \quad (5)$$

will depend on $\mathbf{X}$ under typical logistic regression models.

In this paper, we develop semiparametric tests of H_0 that do not suffer from the above difficulties. In Section 2, this is realized by considering a conditional mean model for the expected counterfactual outcomes, given the confounder $\mathbf{X}$, which is defined by a parametric model for the probability contrast (5), but which is otherwise unrestricted. In Section 3, this is realized by instead focussing on the marginal probability contrast

$$\sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}). \quad (6)$$

where $\pi(\mathbf{a}) = E(Y_{\mathbf{a}})$. Tests of

$$H_0: \sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}) \leq 0 \quad (7)$$

can detect sufficient cause interactions at the population level and will be referred to as tests for marginal sufficient cause interaction, as opposed to tests for conditional sufficient cause interaction based on the $\mathbf{X}$ -specific contrast (5). In contrast to the latter, tests for marginal sufficient cause interaction do not require the existence of a common confounder $\mathbf{X}$ for the effects of both exposures A_1 and A_2 on the outcome, and can in particular be used even when some of the confounders are intermediate (i.e. on the causal path from one of the exposures to the outcome). The performance of the different testing approaches is evaluated through simulation studies in Section 4 and through a data analysis in Section 5.

2 Conditional sufficient cause interaction

2.1 The general case

To estimate the functional (5), we postulate model $\mathcal{A}$ characterized by the sole restriction

$$\sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}) = q(\mathbf{X}; \beta^*), \quad (8)$$

with $\lambda(\mathbf{a})$ a known function, $q(\mathbf{X}; \beta)$ a known function of $\mathbf{X}$, smooth in β and satisfying $q(\mathbf{X}; \beta) = 0$ for all $\mathbf{X}$ if and only if $\beta = 0$, and β^* an unknown, finite-dimensional parameter. A test of the null hypothesis that $q(\mathbf{X}; \beta^*) = 0$ w.p.1 is thus a test of the hypothesis (3).

For instance, to test restriction (2), we might postulate the model $q(\mathbf{X}; \beta^*) = \beta^*$ so that

$$\pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) - \pi((1, 0), \mathbf{X}) + \pi((0, 0), \mathbf{X}) = \beta^*, \quad (9)$$

and test the null $\beta^* = 0$ versus the alternative $\beta^* > 0$. The model defined by this restriction is equivalent to the semiparametric interaction model (Vansteelandt et al., 2008b)

$$E(Y | \mathbf{A}, \mathbf{X}) = \beta^* A_1 A_2 + q_2(\mathbf{X}) A_2 + q_1(\mathbf{X}) A_1 + q_0(\mathbf{X}), \quad (10)$$

which specifies the interaction between both exposures, but leaves the main effects $q_j(\mathbf{X})$, $j = 0, 1, 2$, unspecified. Likewise, to test restriction (1), we might postulate that $q(\mathbf{X}; \beta^*) = \beta^*$ so that

$$\pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) - \pi((1, 0), \mathbf{X}) = \beta^*, \quad (11)$$

and subsequently test $\beta^* = 0$ versus $\beta^* > 0$.

Theorem 1—If $\hat{\beta}$ is a regular asymptotically linear (RAL) estimator of β^* in model $\mathcal{A}$, then there exists a $p \times 1$ function $d(\mathbf{X})$, where p is the dimension of β^* , such that $\hat{\beta}$ has influence function

$$d(\mathbf{X}) \frac{\lambda(\mathbf{A})}{P(\mathbf{A}|\mathbf{X})} \epsilon(\beta), \quad (12)$$

where $\epsilon(\beta) = Y - \pi(\mathbf{A}, \mathbf{X}; \beta)$ and $\pi(\mathbf{A}, \mathbf{X}; \beta)$ is an unknown function, smooth in β and satisfying $\pi(\mathbf{A}, \mathbf{X}; \beta^*) = \pi(\mathbf{A}; \mathbf{X})$ and $\sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}; \beta) = q(\mathbf{X}; \beta)$ for all β . That is,

$$n^{1/2} (\hat{\beta} - \beta^*) = n^{-1/2} \sum_{i=1}^n d(\mathbf{X}_i) \frac{\lambda(\mathbf{A}_i)}{P(\mathbf{A}_i|\mathbf{X}_i)} \epsilon_i(\beta^*) + o_p(1).$$

Remark—We here sketch a proof which is more insightful than the formal proof of the Online Supplement (Section 2). It is immediate that when $\mathbf{A}$ is uniformly distributed and independent of $\mathbf{X}$, the orthocomplement to the nuisance tangent space (Bickel et al., 1993) under model $\mathcal{A}$ is $\{d(\mathbf{X})\lambda(\mathbf{A})\epsilon(\beta^*) : d(\mathbf{X}) \text{ arbitrary}\}$. The result now follows by the Radon-Nikodym theorem.

It follows from Theorem 1 that all regular and asymptotically linear (RAL) estimators of β^* in model $\mathcal{A}$ can be obtained (up to asymptotic equivalence) as the solution $\hat{\beta}(d)$ to the equation

$$\sum_{i=1}^n d(\mathbf{X}_i) \frac{\lambda(\mathbf{A}_i)}{P(\mathbf{A}_i|\mathbf{X}_i)} \epsilon_i(\beta) = 0, \quad (13)$$

for some $d(\mathbf{X}_i)$. For instance, β^* indexing the model defined by (9) can be estimated as the solution to an estimating equation of the form

$$0 = \sum_{i=1}^n d(\mathbf{X}_i) \frac{(-1)^{I(A_{i1} \neq A_{i2})}}{P(\mathbf{A}_i|\mathbf{X}_i)} \epsilon_i(\beta),$$

and β^* indexing the model defined by (11) as the solution to an equation of the form

$$0 = \sum_{i=1}^n d(\mathbf{X}_i) \frac{I(A_{i1}=A_{i2}=1) - I(A_{i1} \neq A_{i2})}{P(\mathbf{A}_i|\mathbf{X}_i)} \epsilon_i(\beta).$$

It further follows from Theorem 1 that the solution $\hat{\beta}(d)$ to equation (13) is an infeasible estimator, as the influence function depends on the unknown conditional law $P(\mathbf{A}_i|\mathbf{X}_i)$ of exposure $\mathbf{A}_i$, given $\mathbf{X}_i$, and $\epsilon_i(\beta)$ depends on the unknown functions $\pi(\mathbf{A}_i, \mathbf{X}_i)$, which cannot be well estimated using smoothing methods in moderate sample sizes when $\mathbf{X}_i$ is a vector with at least several continuous components. As a consequence any feasible estimator of β^* under model $\mathcal{A}$ will exhibit poor finite sample performance in that case. It follows that in

general inference about β^* in model $\mathcal{A}$ is infeasible due to the curse of dimensionality and that dimension-reducing (e.g. parametric) working models must be used.

We consider two possible dimension-reducing strategies. The first is to postulate a parametric model $\mathcal{M}_a$ for the conditional probability of $\mathbf{A}$, given $\mathbf{X}$

$$P(\mathbf{A}|\mathbf{X}) = P(\mathbf{A}|\mathbf{X}; \alpha^*),$$

where $P(\mathbf{A}|\mathbf{X}; \alpha)$ is a known probability function smooth in α and α^* is an unknown finite-dimensional parameter. The second strategy is to postulate a parametric model $\mathcal{M}_y$ for the conditional outcome mean

$$E(Y|\mathbf{A}, \mathbf{X}) = \pi(\mathbf{A}, \mathbf{X}; \beta^*, \eta^*),$$

where we define $\pi(\mathbf{A}, \mathbf{X}; \beta, \eta)$ to be a known function, smooth in η and satisfying the restriction (8) of model $\mathcal{A}$ given by $\sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \pi(\mathbf{A}, \mathbf{X}; \beta, \eta) = q(\mathbf{X}; \beta)$ for all (β, η) , and where η^* is an unknown finite-dimensional parameter. For instance, one may choose arbitrary binary regression models (e.g. logistic regression models) for the probabilities $\pi(\mathbf{a}, \mathbf{X}; \beta, \eta) = \pi(\mathbf{a}, \mathbf{X}; \eta)$ for all $\mathbf{a}$, involving nuisance parameters η only, except for an arbitrary exposure level $\mathbf{a}^*$ for which $\lambda(\mathbf{a}^*) > 0$. To satisfy restriction (8), one may then define

$$\pi(\mathbf{a}^*, \mathbf{X}; \beta, \eta) = \frac{q(\mathbf{X}; \beta)}{\lambda(\mathbf{a}^*)} - \sum_{\mathbf{a} \in \Omega \setminus \{\mathbf{a}^*\}} \pi(\mathbf{a}, \mathbf{X}; \eta) \frac{\lambda(\mathbf{a})}{\lambda(\mathbf{a}^*)}.$$

For instance, in the model defined by (9), a parametric model $\mathcal{M}_y$ can be postulated by specifying arbitrary binary regression models for the probabilities $\pi((1, 0), \mathbf{X})$, $\pi((0, 1), \mathbf{X})$ and $\pi((0, 0), \mathbf{X})$, and next defining

$$\pi((1, 1), \mathbf{X}; \beta, \eta) = \beta + \pi((0, 1), \mathbf{X}; \eta) + \pi((1, 0), \mathbf{X}; \eta) - \pi((0, 0), \mathbf{X}; \eta).$$

Note that this approach is not entirely satisfactory because it is difficult to postulate models which guarantee the probability that $Y = 1$ among subjects with $\mathbf{A} = \mathbf{a}$ to stay between 0 and 1 for all $\mathbf{a}$ and $\mathbf{X}$ and thus the chosen model $\mathcal{M}_y$ risks being logically impossible. In view of this, we will develop estimators $\hat{\beta}(\mathbf{d})$ of β^* that are guaranteed to be consistent and asymptotically normal (CAN) when model $\mathcal{M}_a$ is correct, even when model $\mathcal{M}_y$ is misspecified or logically impossible. As we will demonstrate below, the proposed estimators will, at the same time, be CAN when model $\mathcal{M}_y$ is correct, even when model $\mathcal{M}_a$ is misspecified. That is, the estimators $\hat{\beta}$ will be CAN in the union submodel $\mathcal{A} \cap (\mathcal{M}_a \cup \mathcal{M}_y)$ of model $\mathcal{A}$ that assumes that either $\mathcal{M}_a$ or $\mathcal{M}_y$ is true. In line with Robins and Rotnitzky (2001), we will refer to such estimators as doubly robust.

Part (i) of Theorem 2 below shows that, under mild regularity conditions, the estimators $\hat{\beta} \equiv \hat{\beta}(\mathbf{d})$ are doubly robust, for $\hat{\beta}(\mathbf{d})$ the solution to

$$0 = \sum_{i=1}^n \mathbf{U}_i(\beta, \hat{\eta}(\beta), \hat{\alpha}) = \sum_{i=1}^n \mathbf{d}(\mathbf{X}_i) \frac{\lambda(\mathbf{A}_i)}{\mathbf{P}(\mathbf{A}_i | \mathbf{X}_i; \hat{\alpha})} \epsilon_i(\beta, \hat{\eta}(\beta)) \quad (14)$$

with $\epsilon_i(\beta, \hat{\eta}(\beta)) \equiv Y_i - \pi(\mathbf{A}_i, \mathbf{X}_i; \beta, \hat{\eta}(\beta))$. Here, $\hat{\alpha}$ is the MLE of α^* under model $\mathcal{M}_\alpha$, which satisfies

$$0 = \sum_{i=1}^n \mathbf{H}_i(\hat{\alpha}) \equiv \sum_{i=1}^n \frac{\partial}{\partial \alpha} \ln \mathbf{P}(\mathbf{A}_i | \mathbf{X}_i; \alpha) |_{\alpha=\hat{\alpha}}$$

and $\hat{\eta}(\beta)$ is the semiparametric efficient estimator of η^* under model $\mathcal{A} \cap \mathcal{M}_y$ with β known, which satisfies (Chamberlain, 1987)

$$0 = \sum_{i=1}^n \mathbf{G}_i(\beta, \hat{\eta}(\beta)) \equiv \sum_{i=1}^n \frac{\{\partial \pi(\mathbf{A}_i, \mathbf{X}_i; \eta) / \partial \eta\}_{|\eta=\hat{\eta}(\beta)}}{\pi(\mathbf{A}_i, \mathbf{X}_i; \beta, \hat{\eta}(\beta)) \{1 - \pi(\mathbf{A}_i, \mathbf{X}_i; \beta, \hat{\eta}(\beta))\}} \epsilon_i(\beta, \hat{\eta}(\beta)).$$

Theorem 2—Under regularity conditions we have the following.

(i) $\sqrt{n}(\hat{\beta} - \beta^*)$ is RAL under model $\mathcal{A} \cap (\mathcal{M}_\alpha \cup \mathcal{M}_y)$ with influence function

$$E^{-1} \left\{ \frac{\partial}{\partial \beta} \mathbf{U}_i^*(\beta, \tilde{\eta}(\beta^*), \tilde{\alpha}) |_{\beta=\beta^*} \right\} \mathbf{U}_i^*(\beta^*, \tilde{\eta}(\beta^*), \tilde{\alpha}) \quad (15)$$

and thus converges in distribution to a $N(\mathbf{0}, \Gamma)$, where

$$\Gamma = E \left(\left[E^{-1} \left\{ \frac{\partial}{\partial \beta} \mathbf{U}_i^*(\beta, \tilde{\eta}(\beta^*), \tilde{\alpha}) |_{\beta=\beta^*} \right\} \mathbf{U}_i^*(\beta^*, \tilde{\eta}(\beta^*), \tilde{\alpha}) \right]^{\otimes 2} \right) \quad (16)$$

with $\tilde{\eta}(\beta)$ and $\tilde{\alpha}$ denoting the probability limits of the estimators $\hat{\eta}(\beta)$ and $\hat{\alpha}$, respectively,

$$\begin{aligned} & \mathbf{U}_i^*(\beta, \eta, \alpha) \\ &= \mathbf{U}_i(\beta, \eta, \alpha) - E \left\{ \frac{\partial}{\partial \eta} \mathbf{U}_i(\beta, \eta, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \eta} \mathbf{G}_i(\beta, \eta) \right\} \mathbf{G}_i(\beta, \eta) \\ & \quad - E \left\{ \frac{\partial}{\partial \alpha} \mathbf{U}_i(\beta, \eta, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \alpha} \mathbf{H}_i(\alpha) \right\} \mathbf{H}_i(\alpha) \end{aligned} \quad (17)$$

and where for any matrix $\mathbf{W}$, $\mathbf{W}^{\otimes 2}$ denotes $\mathbf{W}\mathbf{W}'$.

(ii) Furthermore, let $\hat{\beta}(\mathbf{d}, \mathbf{G}_{(1)}, \mathbf{H}_{(1)})$ and $\hat{\beta}(\mathbf{d}, \mathbf{G}_{(2)}, \mathbf{H}_{(2)})$ be 2 estimators of β^* under model $\mathcal{A} \cap (\mathcal{M}_\alpha \cup \mathcal{M}_y)$ corresponding to the same index functions $\mathbf{d}$, but different unbiased estimating functions $\mathbf{G}_{(1)}$ and $\mathbf{G}_{(2)}$ for η^* under model $\mathcal{M}_y$ and $\mathbf{H}_{(1)}$ and $\mathbf{H}_{(2)}$ for α^* under

model $\mathcal{M}_a$. Then, $\sqrt{n} \left\{ \hat{\beta} \left(\mathbf{G}_{(1)}, \mathbf{H}_{(1)} \right) - \hat{\beta} \left(\mathbf{G}_{(2)}, \mathbf{H}_{(2)} \right) \right\} = o_p(1)$ at the intersection submodel $\mathcal{A} \cap \mathcal{M}_a \cap \mathcal{M}_y$.

Theorem 3 shows that a locally semiparametric efficient estimator of β^* in model $\mathcal{A} \cap (\mathcal{M}_y \cup \mathcal{M}_a)$ is obtained by choosing $\mathbf{d}(\mathbf{X}_i)$ in equation (13) equal to

$$\mathbf{d}_{opt}(\mathbf{X}_i) = \frac{\partial q(\mathbf{X}; \beta)}{\partial \beta} \Big|_{\beta=\beta^*} E \left[\pi(\mathbf{A}, \mathbf{X}) \{1 - \pi(\mathbf{A}, \mathbf{X})\} \frac{\lambda(\mathbf{A})^2}{P(\mathbf{A}|\mathbf{X})^2} \Big| \mathbf{X} \right]^{-1}. \quad (18)$$

Theorem 3—(i) The efficient score for β^* in model $\mathcal{A} \cap (\mathcal{M}_y \cup \mathcal{M}_a)$ is $\mathbf{S}_{eff}$, which is given by (12) with $\mathbf{d}(\mathbf{X}_i)$ replaced by $\mathbf{d}_{opt}(\mathbf{X}_i)$.

(ii) Let $\mathbf{d}_{opt}(\mathbf{X}_i; \hat{\alpha}, \hat{\eta}(\beta))$ be defined like $\mathbf{d}_{opt}(\mathbf{X}_i)$, but with $P(\mathbf{A}_i|\mathbf{X}_i)$ replaced by $P(\mathbf{A}_i|\mathbf{X}_i; \hat{\alpha})$ and $\pi(\mathbf{A}_i, \mathbf{X}_i)$ by $\pi(\mathbf{A}_i, \mathbf{X}_i; \hat{\eta}(\beta))$. Then, $\hat{\beta}(\mathbf{d})$ and $\hat{\beta}(\mathbf{d}_{opt}(\hat{\alpha}, \hat{\eta}(\beta)))$ are RAL estimators in model $\mathcal{A} \cap (\mathcal{M}_y \cup \mathcal{M}_a)$. If, in addition, the true distribution of the data lies in the intersection submodel $\mathcal{A} \cap \mathcal{M}_y \cap \mathcal{M}_a$, then the difference between the asymptotic variance matrices of $\hat{\beta}(\mathbf{d})$ and $\hat{\beta}(\mathbf{d}_{opt}(\hat{\alpha}, \hat{\eta}(\beta)))$ is non-negative definite, with the asymptotic variance of $\hat{\beta}(\mathbf{d}_{opt}(\hat{\alpha}, \hat{\eta}(\beta)))$ equalling $\{\text{Var}(\mathbf{S}_{eff})\}^{-1}$.

The approach considered thus far accommodates the concerns about specifying the outcome regression $\mathcal{M}_y$, provided that the joint exposure distribution is known, but nevertheless relies on a model for the probability contrast (5) which can itself be difficult to specify. This raises concerns whether tests of the composite null hypothesis (3) based on the proposed approach may be biased under misspecification of model $\mathcal{A}$. As we will show next, the simple choice $q(\mathbf{X}; \beta) = \beta$ prevents such bias from occurring (provided that the chosen index function $d(\mathbf{X})$ is uniformly non-negative). The proposed test of the null $\beta^* = 0$ versus the alternative $\beta^* > 0$ is then an asymptotically distribution-free test of the null hypothesis of no sufficient cause interaction in settings where the joint exposure distribution is a priori known, as would be the case in either a randomized experiment with all exposures randomly assigned or in a family-based gene-gene interaction study where the law of the genetic markers A_j , $j = 1, \dots, s$ in $\mathbf{A} = (A_1, \dots, A_s)$ is determined by Mendelian inheritance. That this is true can be seen more generally when each component $d(\mathbf{X})$ of $\mathbf{d}(\mathbf{X})$ is non-negative and $q(\mathbf{X}; \beta) = \beta_0 \exp\{k(\mathbf{X}; \beta)\}$ where $k(\mathbf{X}; \beta)$ is an arbitrary function and β_0 is one of the components of β . It then follows from Theorem 1 that the probability limits $\tilde{\beta}$ and $\tilde{\beta}$ satisfy the equation

$$E\{\mathbf{d}(\mathbf{X}) q(\mathbf{X})\} = E\left[\mathbf{d}(\mathbf{X}) \tilde{\beta}_0 \exp\{k(\mathbf{X}; \tilde{\beta})\}\right], \text{ from which}$$

$\tilde{\beta}_0 = E\{d(\mathbf{X}) q(\mathbf{X})\} / E\left[d(\mathbf{X}) \exp\{k(\mathbf{X}; \tilde{\beta})\}\right]$ is non-positive under the null hypothesis, regardless of the choice of $k(\mathbf{X}; \beta)$ and $d(\mathbf{X})$ (with $d(\mathbf{X})$ an arbitrary component of $\mathbf{d}(\mathbf{X})$). By a similar argument one can show that under these conditions, the proposed test of conditional sufficient cause interaction remains consistent (i.e., has power converging to 1 as the sample size approaches infinity) under misspecification of model $\mathcal{A}$ when $q(\mathbf{X}) > 0$ w.p. 1, but not necessarily otherwise. Note from Theorem 3 that the condition of each component

of $\mathbf{d}(\mathbf{X})$ having a uniform sign over all $\mathbf{X}$ is in particular satisfied for the efficient score when $q(\mathbf{X}; \beta)$ is of the form $\beta_0 \exp \{k(\mathbf{X}; \beta)\}$.

When no a priori information is available on the $\mathbf{X}$ -directions in which deviations from the null hypothesis may exist, then the simple choice $q(\mathbf{X}; \beta) = \beta$ is of interest for the aforementioned robustness property and because it avoids the need for $\mathbf{X}$ -specific tests. In contrast, when a priori information is available on the $\mathbf{X}$ -directions in which deviations from the null hypothesis may exist, or there is a specific interest in testing for sufficient cause interactions within subpopulations defined by $\mathbf{X}$, then more general choices of $q(\mathbf{X}; \beta)$ may need to be considered. However, these will imply $\mathbf{X}$ -specific tests which can be practically cumbersome and may be biased as tests of null hypothesis (3) under misspecification of model $\mathcal{A}$. In view of this, we recommend the simple model $q(\mathbf{X}; \beta) = \beta$ for routine data analysis.

2.2 Conditionally independent exposures

Suppose now that $\mathbf{A}$ consists of two univariate components A_1 and A_2 which are conditionally independent, given $\mathbf{X}$, so that $P(\mathbf{A}|\mathbf{X}) = P(A_1|\mathbf{X})P(A_2|\mathbf{X})$. This assumption is often plausible in gene-gene interaction studies of unlinked genetic markers or in gene-environment studies where A_1 and A_2 refer to genetic marker scores and/or environmental exposures. This special case is of interest because, as we show below, it allows for computationally attractive estimators with even greater robustness against model misspecification.

Let $\mathcal{A}_{cip}$ be defined like model $\mathcal{A}$ with the additional assumption that $A_1 \perp\!\!\!\perp A_2|\mathbf{X}$. Then, under model $\mathcal{A}_{cip}$, we can write $\mathcal{M}_a = \mathcal{M}_{a1} \cap \mathcal{M}_{a2}$ where $\mathcal{M}_{aj}, j=1, 2$ is a model for the conditional probability of A_j , given $\mathbf{X}$

$$P(A_j|\mathbf{X}) = P(A_j|\mathbf{X}; \alpha_j^*),$$

where $P(A_j|\mathbf{X}; \alpha_j)$ is a known probability function smooth in α_j and α_1^*, α_2^* are unknown, variation independent, finite-dimensional parameters. Let $\mathcal{M}_{y1}$ be the model for

$$\sum_{a_1} \lambda(a_1, A_2) \pi((a_1, A_2), \mathbf{X}) = \tilde{\pi}(A_2, \mathbf{X}; \beta^*, \eta_1^*),$$

where $\tilde{\pi}(A_2, \mathbf{X}; \beta^*, \eta_1)$ is a known function, smooth in η_1 and satisfying the restrictions of model $\mathcal{A}$, and where η_1^* is an unknown finite-dimensional parameter. Define $\mathcal{M}_{y2}$ like $\mathcal{M}_{y1}$, but with the roles of A_1 and A_2 reversed and with the corresponding parameter η_2 variation independent of η_1 . Define $\mathcal{M}_{y3}$ to be a parametric model for the conditional outcome mean

$$E(Y|\mathbf{A}, \mathbf{X}) = \pi(\mathbf{A}, \mathbf{X}; \beta^*, \eta^*),$$

where we define $\pi(\mathbf{a}, \mathbf{X}; \beta, \boldsymbol{\eta})$ to be a known function, satisfying the restrictions of model $\mathcal{A} \cap \mathcal{M}_{y1} \cap \mathcal{M}_{y2}$ and smooth in $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)'$ with $\boldsymbol{\eta}_3$ variation independent of $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, and where $\boldsymbol{\eta}^* = (\boldsymbol{\eta}_1^*, \boldsymbol{\eta}_2^*, \boldsymbol{\eta}_3^*)'$ with $\boldsymbol{\eta}_3^*$ an unknown finite-dimensional parameter. Finally, define $\mathcal{M}_y = \mathcal{M}_{y1} \cap \mathcal{M}_{y2} \cap \mathcal{M}_{y3}$.

For instance, consider the model defined by (9) or, equivalently, (10). Then postulating $\mathcal{M}_{y1}$ is equivalent to postulating a model for the main effect of A_1 , i.e. $q_1(\mathbf{X}) = q_1(\mathbf{X}; \boldsymbol{\eta}_1^*)$, because

$$\begin{aligned} \sum_{a_1} \lambda(a_1, A_2) \pi((a_1, A_2), \mathbf{X}) &= A_2 \{ \pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) \} + (1 - A_2) \{ \pi((0, 0), \mathbf{X}) - \pi((1, 0), \mathbf{X}) \} \\ &= A_2 \{ \beta^* + q_1(\mathbf{X}) \} - (1 - A_2) q_1(\mathbf{X}) \end{aligned}$$

and, likewise, postulating $\mathcal{M}_{y2}$ is equivalent to postulating a model for the main effect of A_2 , i.e. $q_2(\mathbf{X}) = q_2(\mathbf{X}; \boldsymbol{\eta}_2^*)$. For instance, one could define $\mathcal{M}_{yj}, j=1, 2$ to be the model given by the restriction $q_j(\mathbf{X}; \boldsymbol{\eta}_j^*) = \boldsymbol{\eta}_j^{*'} \mathbf{X}$. It then follows that $\mathcal{M}_{y3}$ is a model for the main effect of $\mathbf{X}$, i.e. $q_0(\mathbf{X}) = q_0(\mathbf{X}; \boldsymbol{\eta}_3^*)$. As a second example, consider the model defined by (11). Then postulating $\mathcal{M}_{y1}$ is equivalent to postulating a model for $\pi((1, 0), \mathbf{X})$ because

$$\begin{aligned} \sum_{a_1} \lambda(a_1, A_2) \pi((a_1, A_2), \mathbf{X}) &= A_2 \{ \pi((1, 1), \mathbf{X}) - \pi((0, 1), \mathbf{X}) \} - (1 - A_2) \pi((1, 0), \mathbf{X}) \\ &= A_2 \{ \beta^* + \pi((1, 0), \mathbf{X}) \} - (1 - A_2) \pi((1, 0), \mathbf{X}) \end{aligned}$$

and likewise postulating $\mathcal{M}_{y2}$ is equivalent to postulating model for $\pi((0, 1), \mathbf{X})$. It thus follows that $\mathcal{M}_{y3}$ is a model for $\pi((0, 0), \mathbf{X})$. For instance, one may choose $\mathcal{M}_{yj}$ for $j = 1, 2, 3$ to be appropriate logistic regression models with variation independent parameters.

With these model definitions, we will now develop estimators $\hat{\beta}_{cip} \equiv \hat{\beta}_{cip}(\mathbf{d})$ of β^* , which are CAN if, in addition to model $\mathcal{A}_{cip}$ either $\mathcal{M}_a$ is correctly specified, or $\mathcal{M}_y$, or $\mathcal{M}_{a1} \cap \mathcal{M}_{y1}$ or $\mathcal{M}_{a2} \cap \mathcal{M}_{y2}$, but not necessarily more than one of them. In line with Vansteelandt et al. (2008b), we will refer to the resulting estimators as quadruply or multiply robust estimators of β^* under model

$$\mathcal{B}_{cip} \equiv \mathcal{A}_{cip} \cap \{ \mathcal{M}_a \cup \mathcal{M}_y \cup (\mathcal{M}_{a1} \cap \mathcal{M}_{y1}) \cup (\mathcal{M}_{a2} \cap \mathcal{M}_{y2}) \}.$$

Theorem 4 below shows that, under mild regularity conditions, the estimator $\hat{\beta}_{cip} \equiv \hat{\beta}_{cip}(\mathbf{d})$ is multiply robust under model $\mathcal{B}_{cip}$ for $\hat{\beta}_{cip}(\mathbf{d})$ the solution to (14) with $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2)'$ and $\hat{\boldsymbol{\alpha}}_j$ satisfying

$$\begin{aligned}
0 &= \sum_{i=1}^n \mathbf{H}_{ij}(\hat{\alpha}_j) \equiv \sum_{i=1}^n \frac{\partial}{\partial \alpha_j} \ln P(\mathbf{A}_{ij} | \mathbf{X}_i; \alpha) |_{\alpha=\hat{\alpha}_j} \\
&\text{for } j=1, 2, \text{ and } \hat{\eta}(\beta) = (\hat{\eta}_1(\beta), \hat{\eta}_2(\beta), \hat{\eta}_3(\beta')) \text{ with } \hat{\eta}_1(\beta) \text{ and } \hat{\eta}_2(\beta) \text{ solving} \\
0 &= \sum_{i=1}^n \mathbf{G}_{ij}(\beta, \hat{\eta}(\beta), \hat{\alpha}_j) \equiv \sum_{i=1}^n \frac{e_j(\mathbf{A}_{ij}', \mathbf{X}_i)^{\lambda(\mathbf{A}_i)}}{P(\mathbf{A}_{ij} | \mathbf{X}_i; \hat{\alpha}_j)} \epsilon_i(\beta, \hat{\eta}(\beta)) \\
&\text{for } j, j'=1, 2 \text{ and } j \neq j' \text{ and } \hat{\eta}_3 \text{ solving} \\
0 &= \sum_{i=1}^n \mathbf{G}_{i3}(\beta, \hat{\eta}(\beta)) \equiv \sum_{i=1}^n e_3(\mathbf{X}_i) \epsilon_i(\beta, \hat{\beta}(\beta)) \\
&\text{for arbitrary vector function } e_j(\mathbf{X}_i), j=1, 2, 3 \text{ of the dimension of } \eta_j.
\end{aligned} \tag{19}$$

Theorem 4—Under regularity conditions and $\beta, \eta_1, \eta_2, \eta_3, \alpha_1$ and α_2 being variation independent, $\sqrt{n}(\hat{\beta}_{cip} - \beta^*)$ is RAL under model $\mathcal{B}_{cip}$ with influence function (15) and thus converges in distribution to a $N(0, \Gamma)$, where Γ is given by (16) with

$$\begin{aligned}
\mathbf{U}_i^*(\beta, \eta, \alpha) &= \mathbf{U}_i(\beta, \eta, \alpha) - E \left\{ \frac{\partial}{\partial \eta} \mathbf{U}_i(\beta, \eta, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \eta} \mathbf{G}_i(\beta, \eta, \alpha) \right\} \mathbf{G}_i(\beta, \eta, \alpha) \\
&- \left[E \left\{ \frac{\partial}{\partial \alpha} \mathbf{U}_i(\beta, \eta, \alpha) \right\} - E \left\{ \frac{\partial}{\partial \eta} \mathbf{U}_i(\beta, \eta, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \eta} \mathbf{G}_i(\beta, \eta, \alpha) \right\} \right. \\
&\quad \times \left. E \left\{ \frac{\partial}{\partial \alpha} \mathbf{G}_i(\beta, \eta, \alpha) \right\} \right] E^{-1} \left\{ \frac{\partial}{\partial \alpha} \mathbf{H}_i(\alpha) \right\} \mathbf{H}_i(\alpha)
\end{aligned} \tag{20}$$

with $\mathbf{H}_i(\alpha) \equiv (\mathbf{H}'_{i1}(\alpha), \mathbf{H}'_{i2}(\alpha))'$ and

$$\mathbf{G}_i(\beta, \eta, \alpha) \equiv (\mathbf{G}'_{i1}(\beta, \eta, \alpha_1), \mathbf{G}'_{i2}(\beta, \eta, \alpha_2), \mathbf{G}'_{i3}(\beta, \eta))'.$$

It can be deduced from the fact that the influence functions for β^* in model $\mathcal{B}_{cip}$ are orthogonal to the nuisance scores for the nonparametric law $P(\mathbf{A}|\mathbf{X})$ under model $\mathcal{M}_a$ that the semiparametric variance bounds in models $\mathcal{A}$ and $\mathcal{A}_{cip}$ are identical whenever model $\mathcal{M}_a$ is true. It can further be deduced from Robins and Rotnitzky (2001) that the semiparametric variance bounds in models $\mathcal{A}_{cip}$ and $\mathcal{B}_{cip}$ are identical whenever the model $\mathcal{M}_y \cap \mathcal{M}_a$ holds. It follows that the choice of index function (18) also yields the semiparametric efficient score under model $\mathcal{B}_{cip}$ at the intersection model $\mathcal{A}_{cip} \cap \mathcal{M}_y \cap \mathcal{M}_a$.

For instance, consider the model defined by (9). Then it is easy to show that for dichotomous exposures A_1 and A_2 , equation (14) is of the form

$$0 = \sum_{i=1}^n d(\mathbf{X}_i) \{ \mathbf{A}_{i1} - P(\mathbf{A}_{i1}=1 | \mathbf{X}_i; \hat{\alpha}_1) \} \{ \mathbf{A}_{i2} - P(\mathbf{A}_{i2}=1 | \mathbf{X}_i; \hat{\alpha}_2) \} \epsilon_i(\beta, \hat{\eta}(\beta)), \tag{21}$$

and that equations (19) are of the form

$$0 = \sum_{i=1}^n e_j(\mathbf{A}_{ij}', \mathbf{X}_i) \{ \mathbf{A}_{ij} - P(\mathbf{A}_{ij}=1 | \mathbf{X}_i; \hat{\alpha}_j) \} \epsilon_i(\beta, \hat{\eta}(\beta))$$

for $j, j' = 1, 2$ and $j \neq j'$. The estimator $\hat{\beta}_{cip}$ is then CAN for β^* when in addition to $\mathcal{A}_{cip}$, either the exposure distribution $\mathcal{M}_a$ is correctly modelled, or the conditional outcome mean $\mathcal{M}_y$, or the main effect of A_1 along with its conditional distribution (i.e. $\mathcal{M}_{a1} \cap \mathcal{M}_{y1}$), or the main effect of A_2 along with its conditional distribution (i.e. $\mathcal{M}_{a2} \cap \mathcal{M}_{y2}$). Furthermore, the estimator obtained by solving (21) with

$$d(\mathbf{X}_i) = \left[\sum_{\mathbf{a} \in \Omega} \{1 - \pi(\mathbf{a}, \mathbf{X}_i)\} \pi(\mathbf{a}, \mathbf{X}_i) | \{a_1 - P(A_{i1}=1|\mathbf{X}_i)\} \{a_2 - P(A_{i2}=1|\mathbf{X}_i)\} | \right]^{-1}$$

is locally efficient under model $\mathcal{B}_{cip}$ at model $\mathcal{A}_{cip} \cap \mathcal{M}_a \cap \mathcal{M}_y$. When the conditional outcome variance, given $\mathbf{A}_i$ and $\mathbf{X}_i$, is constant, then the choice $d(\mathbf{X}_i) = 1$ in equation (21) yields a locally efficient estimator under model $\mathcal{B}_{cip}$, suggesting that this simple choice gives a reasonably efficient estimator; however, simulation studies revealed somewhat better performance with the locally efficient choice. Consider now the model defined by (11). Then it is easy to show that for dichotomous exposures A_1 and A_2 , equation (14) is of the form

$$0 = \sum_{i=1}^n d(\mathbf{X}_i) \frac{\{\mathbf{A}_{i1} - P(\mathbf{A}_{i1}=1|\mathbf{X}_i; \hat{\alpha}_1)\} \{\mathbf{A}_{i2} - P(\mathbf{A}_{i2}=1|\mathbf{X}_i; \hat{\alpha}_2)\}}{1 - P(\mathbf{A}_{i1}=1|\mathbf{X}_i; \hat{\alpha}_1) P(\mathbf{A}_{i2}=1|\mathbf{X}_i; \hat{\alpha}_2)} \times (\mathbf{A}_{i1} + \mathbf{A}_{i2} - \mathbf{A}_{i1} \mathbf{A}_{i2}) \epsilon_i(\beta, \hat{\eta}(\beta)), \quad (22)$$

and that equations (19) are of the form

$$0 = \sum_{i=1}^n e_j(\mathbf{A}_{ij}, \mathbf{X}_i) \{\mathbf{A}_{ij} - P(\mathbf{A}_{ij}=1|\mathbf{X}_i; \hat{\alpha}_j)\} (\mathbf{A}_{i1} + \mathbf{A}_{i2} - \mathbf{A}_{i1} \mathbf{A}_{i2}) \epsilon_i(\beta, \hat{\eta}(\beta))$$

for $j, j' = 1, 2$ and $j \neq j'$. The estimator $\hat{\beta}_{cip}$ is then CAN for β^* when in addition to $\mathcal{A}_{cip}$, either the exposure distribution $\mathcal{M}_a$ is correctly modelled, or the conditional outcome mean $\mathcal{M}_y$, or the conditional outcome mean given $(A_1 = 1, A_2 = 0, \mathbf{X})$ (i.e., $\pi(1, 0, \mathbf{X})$) along with the conditional distribution of A_1 (i.e. $\mathcal{M}_{a1} \cap \mathcal{M}_{y1}$), or the conditional outcome mean given $(A_1 = 0, A_2 = 1, \mathbf{X})$ (i.e., $\pi(0, 1, \mathbf{X})$) along with the conditional distribution of A_2 (i.e. $\mathcal{M}_{a2} \cap \mathcal{M}_{y2}$). Furthermore, the estimator obtained by solving (22) with

$$d(\mathbf{X}_i) = \{1 - P(A_{i1}=1|\mathbf{X}_i) P(A_{i2}=1|\mathbf{X}_i)\} \left[\sum_{\mathbf{a}} \{1 - \pi(\mathbf{a}, \mathbf{X}_i)\} \pi(\mathbf{a}, \mathbf{X}_i) \times (a_1 + a_2 - a_1 a_2) | \{a_1 - P(A_{i1}=1|\mathbf{X}_i)\} \{a_2 - P(A_{i2}=1|\mathbf{X}_i)\} | \right]^{-1}$$

is locally efficient under model $\mathcal{B}_{cip}$ at model $\mathcal{A}_{cip} \cap \mathcal{M}_a \cap \mathcal{M}_y$. When the conditional outcome variance, given $\mathbf{A}_i$ and $\mathbf{X}_i$, is constant, then $d(\mathbf{X}_i) = 1$, suggesting that this choice in equation (22) gives a reasonably efficient estimator. Finally, because the estimators considered in this section have influence functions of the form (13), the properties described at the end of Section 2.1 continue to apply: when $q(\mathbf{X}; \beta) = \beta$, the proposed tests of H_0 are asymptotically distribution-free tests of the null hypothesis of no sufficient cause interaction when the joint exposure distribution is a priori known.

3 Marginal sufficient cause interaction

In the previous section, we remedied the difficulties in testing linear contrasts between conditional probabilities by modeling only the probability contrast of interest. As suggested by VanderWeele, Vansteelandt and Robins (2010), an alternative is to focus instead on the marginal probability contrast (6). Because the model given by (6) equalling an unknown value β^* imposes no restrictions on the observed data law, this gives as an advantage that it is compatible with arbitrary models for $\mathcal{M}_a$ and $\mathcal{M}_y$, in particular with standard logistic outcome regression models. In this section, we will explain how to draw inference for β^* when there is a common confounder $\mathbf{X}$ such that conditioning on $\mathbf{X}$ suffices to control for confounding for the estimation of the causal effect of $\mathbf{A}$ on $\mathbf{Y}$ (in the sense that $Y_a \perp\!\!\!\perp \mathbf{A}|\mathbf{X}$ for all $\mathbf{a} \in \Psi$). This assumption may not be realistic, for instance, in the context of a gene-environment interaction study where some of the confounders of the association between the environmental exposure and the outcome are themselves causally affected by the studied gene (VanderWeele, Vansteelandt and Robins, 2010). In such settings, adjustment for these confounders could induce a bias in the estimates of the genetic effect and thus also in the counterfactual expectations $\pi(\mathbf{a}, \mathbf{X})$. In Section 4 of the Online Supplement, we show how the here proposed tests for marginal sufficient cause interaction can be adapted in that case.

VanderWeele, Vansteelandt and Robins (2010) propose a simple regression approach to estimating β^* under model $\mathcal{M}_a$, which more generally amounts to calculating

$$\hat{\beta}_{IPW} = \sum_{\mathbf{a} \in \Omega} \lambda(\mathbf{a}) \frac{\sum_{i=1}^n I(\mathbf{A}_i = \mathbf{a}) Y_i / P(\mathbf{A}_i | \mathbf{X}_i; \hat{\alpha})}{\sum_{i=1}^n I(\mathbf{A}_i = \mathbf{a}) / P(\mathbf{A}_i | \mathbf{X}_i; \hat{\alpha})}. \quad (23)$$

We here propose a doubly robust estimator for β^* under the union model $\mathcal{M}_a \cup \mathcal{M}_y$, where we redefine $\mathcal{M}_y$ to be a model for the conditional outcome mean $E(Y|\mathbf{A}, \mathbf{X}) = \pi(\mathbf{A}, \mathbf{X}; \boldsymbol{\eta}^*)$, with $\pi(\mathbf{A}, \mathbf{X}; \boldsymbol{\eta})$ a known function, smooth in $\boldsymbol{\eta}$, and with $\boldsymbol{\eta}^*$ an unknown finite-dimensional parameter.

Theorem 5 in Section 4 of the Online Supplement shows that the estimator

$$\hat{\beta}_{mar} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\lambda(\mathbf{A}_i) \{Y_i - \pi(\mathbf{A}_i, \mathbf{X}_i; \hat{\boldsymbol{\eta}})\}}{P(\mathbf{A}_i | \mathbf{X}_i; \hat{\alpha})} + \sum_{\mathbf{a}} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}_i; \hat{\boldsymbol{\eta}}) \right]. \quad (24)$$

is a CAN estimator under model $\mathcal{M}_a \cup \mathcal{M}_y$, where $\hat{\alpha}$ is as defined previously and $\hat{\boldsymbol{\eta}}$ is a CAN estimator of $\boldsymbol{\eta}$ under model $\mathcal{M}_y$. The latter would most naturally be obtained through maximum likelihood estimation under model $\mathcal{M}_y$, although we here recommend obtaining it as the solution to an estimating equation which imposes

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{\lambda(\mathbf{A}_i) \{Y_i - \pi(\mathbf{A}_i, \mathbf{X}_i; \boldsymbol{\eta})\}}{P(\mathbf{A}_i | \mathbf{X}_i; \hat{\alpha})},$$

so that

$$\hat{\beta}_{mar} = \frac{1}{n} \sum_{i=1}^n \sum_{\mathbf{a}} \lambda(\mathbf{a}) \pi(\mathbf{a}, \mathbf{X}_i; \hat{\boldsymbol{\eta}}).$$

For instance, when $\mathbf{A} = (A_1, A_2)'$ with A_1 and A_2 dichotomous, then this is satisfied when $\pi(\mathbf{A}, \mathbf{X}; \boldsymbol{\eta})$ is a logistic regression model with intercept, main effects and interaction between A_1 and A_2 , and when $\hat{\boldsymbol{\eta}}$ is obtained as the solution to the weighted score equations with weights $1/P(\mathbf{A}_i|\mathbf{X}_i; \hat{\boldsymbol{\alpha}})$. Similar suggestions were made by Robins et al. (2007), who referred to the resulting doubly robust estimators as ‘regression doubly robust estimators’. They have been found to insulate doubly robust estimators against inflation of model misspecification bias in regions of the covariate space where the probabilities $P(\mathbf{A}_i|\mathbf{X}_i; \hat{\boldsymbol{\alpha}})$ are small (Vansteelandt, Bekaert and Claeskens, 2010). Let $\mathcal{A}_{mar,cip}$ be the model defined by the assumption that $\mathbf{A}$ consists of 2 univariate components A_1 and A_2 , which are conditionally independent, given $\mathbf{X}$. Then we show in Section 5 of the Online Supplement that $\hat{\beta}_{mar}$ is also a CAN estimator of β^* in the less restrictive model $\mathcal{A}_{mar,cip} \cap \{\mathcal{M}_q \cup \mathcal{M}_y \cup (\mathcal{M}_{a1} \cap \mathcal{M}_{y1}) \cup (\mathcal{M}_{a2} \cap \mathcal{M}_{y2})\}$, provided that $\hat{\boldsymbol{\eta}}$ is obtained as in Theorem 4.

4 Simulation study

We conducted several series of simulation experiments to evaluate and compare the performance of the different estimators. In a first series, in each of 5000 simulation experiments, each with sample size 1000, we generated a standard normal variate X , conditionally independent Bernoulli variates A_1 and A_2 with $P(A_j = 1/X) = \text{expit}(-1 + \alpha X)$ for $j = 1, 2$, where $\text{expit}(x) = \exp(x)/\{1 + \exp(x)\}$. We next generated Y_{ab} for $a, b = 0, 1$ as independent Bernoulli variates with $P(Y_{00} = 1/X) = \text{expit}(\eta_0 + \eta_X X)$, $P(Y_{10} = 1|Y_{00} = 0, X) = \text{expit}(\eta_0 + \eta_1 + \eta_X^* X)$, $P(Y_{01} = 1|Y_{00} = 0, X) = \text{expit}(\eta_0 + \eta_2 + \eta_X X)$ and chose $P(Y_{11} = 1|Y_{10} = Y_{01} = 0, X)$ such that $P(Y_{11} = 1/X) = P(Y_{10} = 1/X) + P(Y_{01} = 1/X) - P(Y_{00} = 1/X) + \beta^*$ with $\beta^* = 0$ at the null hypothesis and $\beta^* = 0.05$ otherwise. In Section 6 of the Online Supplement, we show that these models are consistent with a sufficient cause representation that satisfies monotonicity. We then evaluated the following estimators of β^* :

1. OLS: A standard linear model analysis with working model

$$E(Y|A_1, A_2, X) = \eta_0 + \eta_1 A_1 + \eta_2 A_2 + \beta A_1 A_2 + \eta_3 X. \quad (25)$$

2. Conditional: The locally efficient estimator obtained by solving (13) with $q(\mathbf{X}; \beta^*) = \beta^*$ and index function as given in (18), using correctly specified working models for the exposure and working model (25) for the outcome.
3. IPW: The estimator (23).
4. Marginal: The doubly robust estimator (24).

In the latter two estimators, inverse weights were truncated at the 99% percentile.

In simulation experiment 1, we chose $\alpha = 2$, $\eta_0 = -5$, $\eta_1 = 0.5$, $\eta_2 = 1$ and $\eta_x = \eta_x^* = 0$. The covariate X is not associated with the outcome in this simulation experiment so that a standard linear model analysis is valid here. In simulation experiment 3, we chose $\eta_x = \eta_x^* = 0.5$, implying that the model $q(X; \beta) = \beta$ is misspecified. In simulation experiment 5, we chose η_x^* twice as large as $\eta_x = 0.5$; this induces an interaction between A_1 and X and hence aggravates the degree of model misspecification. In these simulation experiments, the weights $1/P(A/X)$ had 1%, 95% and 99% percentiles equalling 1.01, 10.3 and 28.6, respectively. Simulation experiments 2, 4 and 6 were like experiments 1, 3 and 5, except that $\alpha = 1$, which led to improved weight stability corresponding to 1%, 95% and 99% percentiles equalling 1.07, 7.6 and 13.7, respectively. Simulations 7-12 and 13-18 were like experiments 1-6, except that $\eta_0 = -4$ and $\eta_0 = -1.25$, respectively, in order to evaluate the impact of a higher outcome prevalence. To avoid values of $P(Y_{11} = 1/X)$ exceeding 1, weaker associations with X were generated, with $\eta_x = 0.35$ in simulations 7-12 and $\eta_x = 0.1$ in simulations 13-18. All simulation results are reported in terms of the empirical bias, the empirical standard deviation and the average sandwich standard error of the estimators for β^* , as well as the Type I error rate or power, which here refer to the test of the null $\beta^* = 0$ versus the alternative $\beta^* \neq 0$. Note that these do not correspond to the Type I error rate or power of the test of the composite null hypothesis (3), but will portray more clearly the impact of model misspecification.

The simulation results in Table 1 were obtained under the null hypothesis ($\beta^* = 0$). They reveal that the OLS estimator can be severely biased under model misspecification (i.e. rows 3-6 and 9-12), leading to an inflation of the Type I error rate. This is less pronounced in rows 15-18 because the weaker associations between X and Y imply a less serious degree of model misspecification. In contrast, the three procedures for sufficient cause interaction yield approximately unbiased estimates. As predicted by the theory, this is so even for the 'Conditional' estimator under misspecification of $q(X; \beta)$. Overall the best results were obtained with the locally efficient 'Conditional' estimator. The simple IPW estimator and the doubly-robust procedure for marginal sufficient cause interaction yielded slightly more biased estimates in all simulation experiments. The doubly robust estimator was usually slightly less precise, but also less biased than the simple IPW estimator. The remaining finite-sample bias of both estimators causes their Type I error rate to deviate slightly from the nominal 5% level. Some conservativeness was also seen for all procedures at low outcome means as a result of the limited information. For the conditional procedure, this can be seen by the fact that the nominal rates are attained at higher outcome means. In the doubly robust estimator of marginal sufficient cause interaction, fitting model $\mathcal{M}_y$ instead using maximum likelihood estimation resulted in empirical standard deviations of the interaction parameters being 15% to 400% larger. Alternative attempts to improve efficiency along the lines of Rubin and van der Laan (2008), Tan (2008) and Cao, Tsiatis and Davidian (2009) resulted in a more minor efficiency loss, but no efficiency gain.

Table 2 evaluates the power of the different procedures at $\beta^* = 0.05$. It can be deduced from rows 1, 2, 7, 8, 13 and 14 (where the linear model (25) holds) that the price to pay for greater robustness against model misspecification is a reduction in power. Note, however, that these discrepancies are partly due to the OLS procedure being too liberal (see Table 1)

as a result of the outcome being non-normal. The latter is consistent with the fact that less pronounced discrepancies are observed when the degree of conservatism diminishes (see rows 13 and 14 of Table 2).

Table 3 shows similar results as in Table 1, but for settings with high-dimensional X . Specifically, to enhance comparability with the previous data-generating mechanism, X was substituted by $\sqrt{10}$ times the average of 10 independent standard normal variates, but the effects of these 10 variates were separately modeled through main effects in the outcome and exposure working models. The results are similar as before, but show that the ordinary least squares estimator is often less efficient than the estimator obtained via the conditional procedure, with biased standard errors. This is because efficiency of the ordinary least squares estimators and unbiasedness of the common likelihood-based standard errors is not guaranteed when the outcome is non-normal.

In a second series of simulation experiments, we evaluated the performance of the tests for sufficient cause interaction without assuming monotonicity under the same data-generating models, but with $P(Y_{00} = 1/X) = 0$ (see Tables 4 and 5). This case is of interest because it implies that the empirical conditions for sufficient cause interactions with and without the assumption of monotonicity are the same (i.e., conditions (1) and (2) coincide in this case). Thus these simulation experiments allow for comparing the power of both corresponding types of tests. Because the empirical conditions for sufficient cause interactions with and without monotonicity are the same, also the corresponding marginal tests for sufficient cause interaction are mathematically equivalent. This is because the underlying models impose the same restrictions on the observed data law and because all subjects with $A_1 = A_2 = 0$ have $Y = 0$. In contrast, the tests for conditional sufficient cause interaction with and without monotonicity are different, with the latter being slightly less powerful.

In a final series of simulation experiments, we generated non-monotonic effects by randomly drawing Y_{ab} for $a, b = 0, 1$ as independent Bernoulli variates with $P(Y_{00} = 1/X) = \text{expit}(\eta_0 + \eta_x X)$, $P(Y_{10} = 1/X) = \text{expit}(\eta_0 + \eta_1 + \eta_x^* X)$, $P(Y_{01} = 1/Y_{00} = 1, X) = 1$, and $P(Y_{01} = 1/Y_{00} = 0, X) = \text{expit}(\eta_0 + \eta_2 + \eta_x X)$. Further, we chose $P(Y_{11} = 1/Y_{00} = Y_{10} = Y_{01} = 0, X)$ equal to

$$\frac{P(Y_{00}=1|X) \{1 - P(Y_{01}=1|Y_{00}=0, X)\} + P(Y_{10}=1|X) P(Y_{01}=1|Y_{00}=0, X)}{1 - P(Y_{10}=1|X) \{1 - P(Y_{01}=1|Y_{00}=0, X)\} + P(Y_{10}=1|X) P(Y_{01}=1|Y_{00}=0, X)}$$

and $P(Y_{11} = 1/Y_{00} = Y_{01} = 1, Y_{10} = 0, X)$ equal to

$$P(Y_{11}=1|Y_{00}=Y_{10}=Y_{01}=0, X) \{1 - P(Y_{01}=1|Y_{00}=0, X)\} + P(Y_{01}=1|Y_{00}=0, X).$$

In Section 6 of the Online Supplement, we motivate this choice of simulation models and demonstrate that $\beta^* = 0$ under these models. Parameter values were chosen as in the first series of simulation experiments.

Table 6 compares three tests for sufficient cause interaction which allow for non-monotonic effects: a test based on ordinary least squares regression (OLS), the test for conditional sufficient cause interaction (Cond. Non-Monot.) and the test for marginal sufficient cause interaction (Marginal Non-Monot.). In addition, we display results for the corresponding tests which assume monotonic effects (Cond. Monot. and Marginal Monot., respectively). The results reveal that incorrectly assuming monotonic effects can yield a sizeable inflation of the Type I error, and that the ordinary least squares analysis can be severely biased under model misspecification. Both proposed tests for sufficient cause interaction which allow for non-monotonic effects perform well, with the conditional estimator having the smallest standard errors and thus most power.

5 Application

We reanalyze a cross-sectional study of 11 062 individuals in Bangladesh (Ahsan et al., 2006; Chen et al., 2006) concerning interactions between the effects of exposure to high levels of arsenic in drinking water (> 150 lg/L in well water) and current or past tobacco smoking on premalignant skin lesions. The example is included for illustrative purposes. Exposure working models were constructed using logistic regression conditional on the covariates sex, age, education, BMI, land and TV ownership, fertilizer use and pesticide use. The monotonicity assumption that smoking and arsenic never prevent skin lesions is biologically plausible here. Under this assumption, we estimate β^* in (9) to equal 0.064 (SE 0.022) using the methods of Section 2.2, thus indicating strong evidence of conditional sufficient cause interaction (P 0.0037). The analysis of marginal sufficient cause interactions yields weaker effects of 0.036 (SE 0.018, P 0.042) using simple IPW and 0.048 (SE 0.026, P 0.069) using doubly robust IPW. As in VanderWeele, Vansteelandt and Robins (2010), there was no evidence of sufficient cause interaction when the monotonicity assumption was dropped because the magnitude of $E(Y_{00})$ exceeds the interaction term.

6 Conclusion

In conclusion, because serious misspecification of linear models for a dichotomous outcome can be logically anticipated, we recommend that tests for sufficient cause interaction (or, more generally tests on the sign of contrasts of outcome probabilities) be based on the locally efficient procedure for conditional sufficient cause interaction with $q(\mathbf{X}; \beta^*) = \beta^*$. This simple model choice is generally recommended (unless there were an interest in testing within subgroups) because it overcomes the need for repeated $\mathbf{X}$ -specific tests and because its misspecification does not affect the test's nominal Type I error rate. The conditional procedure is generally recommended because it outperforms procedures targeting marginal sufficient cause interaction. The proposed procedures for marginal sufficient cause interaction are recommended when intermediate confounding is present and whenever there is an additional interest in the magnitude of the probability contrast (6) (i.e., when the focus is on estimation rather than testing). The latter would be of interest when measuring the prevalence of sufficient cause interactions (VanderWeele, Vansteelandt and Robins, 2010).

An attractive feature of the proposed approach is that it provides an asymptotically distribution-free test of the null hypothesis of no sufficient cause interaction when a priori

information regarding the joint exposure distribution is known. Often such information is not known; in that case, the proposed procedure remains attractive because it offers the analyst multiple chances instead of one to obtain valid results: when either a correct model for the exposure distribution $\mathcal{M}_a$ or for the expected outcome mean $\mathcal{M}_y$ has been postulated. While correct specification of $\mathcal{M}_y$ in a way that is compatible with the semiparametric model $\mathcal{A}$ is not likely when testing for conditional sufficient cause interaction, this double robustness property of our estimators ensures that, provided the sample size is sufficiently large, they will have little bias so long as model $\mathcal{M}_y$ is not grossly misspecified, even when $\mathcal{M}_a$ is incorrect. For instance, it is easily shown using similar arguments as in Vansteelandt, Bekaert and Claeskens (2010) that the asymptotic bias of the solution to (21) is

$$\frac{E \{d(\mathbf{X}_i) \Delta_1(\mathbf{X}_i) \Delta_2(\mathbf{X}_i) \Delta(\mathbf{X}_i)\}}{E \{d(\mathbf{X}_i) P(A_{i1}=1|\mathbf{X}_i) P(A_{i1}=0|\mathbf{X}_i; \tilde{\alpha}_1) P(A_{i2}=1|\mathbf{X}_i) P(A_{i2}=0|\mathbf{X}_i; \tilde{\alpha}_2)\}},$$

where $\Delta_j(\mathbf{X}_i) = P(A_{ij}=1|\mathbf{X}_i) - P(A_{ij}=1|\mathbf{X}_i; \tilde{\alpha}_j)$, $j = 1, 2$, measures conditional bias in the exposure models and likewise $\Delta(\mathbf{X}_i)$ measures conditional bias in the outcome model. This expression shows that any bias due to misspecification in either the exposure or outcome working model gets inflated only by the amount of misspecification affecting the other working models. The same is not generally true for doubly robust estimators of marginal sufficient cause interaction. As suggested by the first term of expression (24), and in line with the findings of Kang and Schafer (2007), any bias affecting the outcome working model can get substantially inflated in regions where the exposure probabilities are small. To avoid this from happening, we have recommended using regression doubly-robust estimators (Robins et al., 2007), which force that first term to be zero (Vansteelandt, Bekaert and Claeskens, 2010).

Efficient and robust methods for inference for sufficient cause interactions have until now been lacking from the literature. The problem of testing linear contrasts of probabilities with covariates constitutes a challenging and important problem in its own right. We have provided a very general and powerful methodology for drawing inferences in these settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The first author acknowledges support from Ghent University (Multidisciplinary Research Partnership Bioinformatics: from nucleotides to networks) and IAP research network grant nr. P06/03 from the Belgian government (Belgian Science Policy).

References

- Ahsan H, Chen Y, Parvez F, et al. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J. Expo. Sci. Environ. Epidemiol.* 2006; 16:191–205. [PubMed: 16160703]
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and adaptive estimation for semiparametric models. Springer-Verlag; New-York: 1993.

- Cao W, Tsiatis AA, Davidian M. Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*. 2009; 96:723–734. [PubMed: 20161511]
- Chamberlain G. Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics*. 1987; 34:305–334.
- Chen Y, Graziano JH, Parvez F, et al. Modification of risk of arsenic-induced skin lesions by sunlight exposure, smoking, and occupational exposures in Bangladesh. *Epidemiology*. 2006; 17:459–467. [PubMed: 16755266]
- Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molec. Genet.* 2002; 11:2463–2468. [PubMed: 12351582]
- Greenland S. Basic problems in interaction assessment. *Envir. Health Persp.* 1993; 101:59–66.
- Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am. J. Epidemiol.* 2004; 160:301–305. [PubMed: 15286014]
- Greenland S. Interactions in Epidemiology: Relevance, Identification, and Estimation. *Epidemiol.* 2009; 20:14–17.
- Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Science*. 2007; 22:523–539.
- Koopman JS. Interaction between discrete causes. *Am. J. Epidemiol.* 1981; 113:716–724. [PubMed: 7234861]
- Maity A, Carroll RJ, Mammen E, Chatterjee N. Testing in semi-parametric models with interaction, with applications to gene-environment interactions. *J. Roy. Statist. Soc. B*. 2009; 71:75–96.
- Mantel N, Brown C, Byar DP. Tests for homogeneity of effect in an epidemiologic investigation. *Am. J. Epidemiol.* 1977; 106:125–129. [PubMed: 888814]
- Murcray CE, Lewinger JP, Gauderman WJ. Gene-Environment Interaction in Genome-Wide Association Studies. *Am. J. Epidemiol.* 2009; 169:219–226.
- Phillips PC. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 2008; 9:855–867. [PubMed: 18852697]
- Robins JM, Rotnitzky A. Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer". *Statistica Sinica*. 2001; 11:920–936.
- Robins JM, Sued M, Lei-Gomez Q, Rotnitzky A. Comment: Performance of double-robust estimators when "Inverse Probability" weights are highly variable. *Statist. Science*. 2007; 22:544–559.
- Rothman KJ. Causes. *Am. J. Epidemiol.* 1976; 104:587–592. [PubMed: 998606]
- Rubin DB, van der Laan MJ. Empirical Efficiency Maximization: Improved Locally Efficient Covariate Adjustment in Randomized Experiments and Survival Analysis. *Int. J. Biostatist.* 2008; 4 Article 5.
- Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. *Am. J. Epidemiol.* 2003; 158:251–258.
- Tan ZQ. A distributional approach for causal inference using propensity scores. *J. Am. Statist. Assoc.* 2006; 101:1619–1637.
- Tan Z. Comment: Improved Local Efficiency and Double Robustness. *Int. J. Bio statist.* 2008; 4 Article 10.
- Tchetgen ET, Robins JM. The semiparametric case-only estimator. *Biometrics*. 2010; 66:1138–1144. [PubMed: 20337632]
- VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiol.* 2009; 20:6–13.
- VanderWeele TJ. Empirical tests for compositional epistasis. *Nat. Rev. Genet.* 2010a; 11:166. [PubMed: 20084088]
- VanderWeele TJ. Epistatic interactions. *Statist. Appl. Genet. Molec. Biol.* 2010b; 9:1–22. Article 1.
- VanderWeele TJ. Sufficient cause interactions for categorical and ordinal exposures with three levels. *Biometrika*. 2010c; 97:647–659. [PubMed: 22822251]
- VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component cause framework. *Epidemiol.* 2007; 18:329–339.
- VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*. 2008; 95:49–61.

- VanderWeele TJ, Richardson TS. General theory for sufficient cause interactions for dichotomous exposures. *Annals of Statistics*, conditionally accepted. 2009
- VanderWeele TJ, Vansteelandt S, Robins JM. Marginal structural models for sufficient cause interactions. *Am. J. Epidemiol.* 2010; 171:506–514.
- Vansteelandt S, DeMeo D, Su J, Smoller J, Murphy AJ, McQueen M, Celedon J, Weiss ST, Silverman EK, Lange C. Testing and estimating gene-environment interactions in family-based association studies. *Biometrics*. 2008a; 64:458–467. [PubMed: 17970814]
- Vansteelandt S, VanderWeele TJ, Tchetgen EJ, Robins JM. Multiply robust inference for statistical interactions. *J. Am. Statist. Assoc.* 2008b; 103:1693–1704.
- Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. *Statist. Meth. Med. Res.* 2010 doi: 10.1177/0962280210387717.

Table 1

Bias, empirical standard deviation, average sandwich standard error and Type I error rate α under the null hypothesis ($\beta^* = 0$) when the exposure effects are monotonic.

	OLS						IPW						Conditional						Marginal					
	$E(Y)$ in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %			
1	1.7	-0.04	2.2	1.8	7.5	-0.09	3.1	2.8	4.1	-0.044	2.7	2.7	2.7	4.4	0.22	3.7	3.3	2.6	4.4	0.22	3.7	3.3	2.6	
2	1.6	-0.03	2.3	1.8	9.8	-0.04	2.8	2.7	2.7	-0.017	2.4	2.4	2.4	4.1	0.11	3.0	2.8	1.3	4.1	0.11	3.0	2.8	1.3	
3	2.4	1.96	2.6	2.2	27.3	0.59	3.2	3.0	8.0	0.036	3.0	3.0	3.0	4.2	0.17	3.2	3.0	4.0	4.2	0.17	3.2	3.0	4.0	
4	2.0	1.56	2.7	2.0	26.2	0.54	2.8	2.7	4.5	0.127	2.8	2.8	2.8	4.3	0.08	2.7	2.6	2.3	4.3	0.08	2.7	2.6	2.3	
5	2.5	3.20	2.6	2.2	43.8	0.79	3.0	2.8	10.0	0.026	2.8	2.8	2.8	3.7	0.20	3.0	2.7	5.3	3.7	0.20	3.0	2.7	5.3	
6	2.1	2.57	2.9	2.0	39.4	0.64	2.6	2.5	5.5	0.174	2.6	2.6	2.6	4.0	0.06	2.5	2.4	2.9	4.0	0.06	2.5	2.4	2.9	
7	4.5	-0.03	3.4	3.0	7.4	-0.01	4.9	4.7	4.5	0.001	4.4	4.4	4.4	4.3	0.27	5.2	5.0	3.2	4.3	0.27	5.2	5.0	3.2	
8	4.1	0.01	3.6	2.9	9.4	0.00	4.4	4.4	3.6	0.053	4.0	4.0	4.0	4.8	0.14	4.6	4.5	2.8	4.8	0.14	4.6	4.5	2.8	
9	5.5	2.89	3.8	3.2	27.3	1.04	5.0	4.8	7.4	0.065	4.7	4.7	4.7	4.6	0.23	4.9	4.7	3.7	4.6	0.23	4.9	4.7	3.7	
10	4.8	2.34	4.0	3.0	25.4	0.99	4.4	4.3	5.8	0.132	4.3	4.3	4.3	4.5	0.13	4.2	4.2	3.2	4.5	0.13	4.2	4.2	3.2	
11	5.2	4.17	3.7	3.1	39.7	1.29	4.6	4.4	9.1	0.077	4.5	4.5	4.5	4.3	0.33	4.5	4.3	4.8	4.3	0.33	4.5	4.3	4.8	
12	5.2	3.80	4.2	3.2	37.3	1.31	4.4	4.4	7.3	0.134	4.4	4.4	4.4	4.6	0.14	4.2	4.2	3.3	4.6	0.14	4.2	4.2	3.3	
13	44.9	0.03	7.0	6.6	5.9	-0.02	9.5	9.2	6.1	-0.041	9.0	9.0	9.0	5.0	-0.10	9.5	9.1	5.8	5.0	-0.10	9.5	9.1	5.8	
14	42.4	-0.03	6.8	6.7	5.1	0.03	8.1	7.9	6.2	-0.030	7.6	7.6	7.7	5.2	-0.02	8.0	7.9	6.0	5.2	-0.02	8.0	7.9	6.0	
15	39.1	2.01	7.0	6.4	10.7	1.25	9.7	9.4	7.7	0.158	9.2	9.2	9.2	5.2	0.84	9.7	9.4	7.2	5.2	0.84	9.7	9.4	7.2	
16	36.6	1.57	6.8	6.5	8.9	1.31	8.4	8.3	7.4	0.049	7.9	7.9	8.0	5.0	0.49	8.4	8.3	6.3	5.0	0.49	8.4	8.3	6.3	
17	38.1	3.31	7.0	6.5	14.8	1.83	9.8	9.5	8.3	0.143	9.3	9.3	9.3	5.1	1.25	9.8	9.4	7.4	5.1	1.25	9.8	9.4	7.4	
18	36.9	2.67	6.8	6.5	11.7	1.79	8.4	8.3	8.3	0.012	7.9	7.9	8.0	5.2	0.77	8.4	8.3	6.9	5.2	0.77	8.4	8.3	6.9	

Table 2
Bias, empirical standard deviation, average sandwich standard error and power under the alternative hypothesis ($\beta^* = 0.05$) when the exposure effects are monotonic.

	E(Y) in %	OLS				IPW				Conditional				Marginal			
		Bias × 100	E.SE × 100	S.S.E × 100	Power in %	Bias × 100	E.SE × 100	S.S.E × 100	Power in %	Bias × 100	E.SE × 100	S.S.E × 100	Power in %	Bias × 100	E.SE × 100	S.S.E × 100	Power in %
1	2.7	-0.08	2.6	2.3	68.4	-0.13	4.0	3.7	38.3	-0.193	3.3	3.3	42.3	0.39	5.0	4.5	25.7
2	2.2	-0.06	2.9	2.1	70.3	-0.05	3.7	3.6	36.3	-0.164	3.2	3.2	45.4	0.18	4.2	4.0	28.0
3	3.4	1.93	2.9	2.6	83.0	0.56	4.0	3.8	44.2	-0.139	3.7	3.7	35.8	-0.23	4.5	4.1	26.3
4	2.7	1.53	3.3	2.3	80.6	0.52	3.7	3.6	42.9	-0.066	3.5	3.6	37.9	-0.25	3.8	3.7	29.6
5	3.5	3.16	2.9	2.6	91.5	0.76	3.9	3.7	50.2	-0.142	3.6	3.6	37.6	-0.37	4.2	3.8	30.2
6	2.7	2.54	3.4	2.3	87.2	0.63	3.6	3.5	47.2	-0.022	3.5	3.5	39.5	-0.38	3.6	3.5	32.4
7	5.5	-0.06	3.7	3.2	46.9	-0.06	5.4	5.2	24.2	-0.094	4.8	4.8	26.4	0.26	5.9	5.6	19.2
8	4.7	-0.03	4.0	3.0	49.9	-0.02	5.0	4.9	23.4	-0.039	4.4	4.4	28.9	0.14	5.2	5.2	20.5
9	6.5	2.87	4.1	3.4	71.2	1.00	5.5	5.3	31.1	-0.015	5.2	5.1	23.9	-0.18	5.6	5.3	19.0
10	5.9	3.75	4.5	3.3	77.1	1.38	5.0	4.9	33.7	0.060	4.9	5.0	25.9	-0.30	4.8	4.8	21.6
11	6.3	4.15	4.0	3.4	82.5	1.26	5.2	5.0	35.1	-0.011	5.0	4.9	25.5	-0.19	5.3	5.0	21.9
12	5.2	3.28	4.3	3.1	77.2	1.11	4.7	4.6	35.0	0.043	4.6	4.6	27.8	-0.25	4.6	4.5	23.3
13	45.9	0.00	6.9	6.5	20.6	-0.04	9.2	8.9	16.0	-0.073	8.7	8.7	14.8	-0.11	9.3	8.9	15.4
14	43.0	-0.04	6.6	6.7	18.3	0.04	7.8	7.7	17.9	-0.036	7.4	7.4	17.5	0.00	7.7	7.6	18.1
15	40.2	1.96	6.9	6.3	31.1	1.13	9.6	9.3	17.7	0.040	9.1	9.0	14.5	0.82	9.6	9.3	17.0
16	37.2	1.57	6.7	6.5	26.7	1.30	8.2	8.1	20.9	0.043	7.8	7.8	16.6	0.55	8.2	8.1	18.9
17	39.2	3.27	6.9	6.4	37.5	1.76	9.7	9.4	19.3	0.080	9.2	9.2	14.5	1.28	9.7	9.3	18.1
18	37.5	2.90	6.6	6.5	33.5	1.96	8.2	8.1	23.2	0.237	7.9	7.8	17.3	1.05	8.3	8.1	21.0

Table 3

Bias, empirical standard deviation, average sandwich standard error and Type I error rate α under the null hypothesis ($\beta^* = 0$) when the exposure effects are monotonic and the confounder is 10-dimensional.

	OLS						IPW						Conditional						Marginal					
	$E(Y)$ in %	Bias $\times 100$	E.S.E. $\times 100$	S.S.E. $\times 100$	α in %	Bias $\times 100$	E.S.E. $\times 100$	S.S.E. $\times 100$	α in %	Bias $\times 100$	E.S.E. $\times 100$	S.S.E. $\times 100$	α in %	Bias $\times 100$	E.S.E. $\times 100$	S.S.E. $\times 100$	α in %	Bias $\times 100$	E.S.E. $\times 100$	S.S.E. $\times 100$	α in %			
1	1.7	0.02	2.2	1.9	8.3	0.04	3.1	2.9	4.6	0.110	2.4	2.2	5.3	0.25	3.4	2.7	6.3							
2	1.5	0.02	2.3	1.8	10.3	0.04	2.8	2.7	2.8	0.083	2.2	2.0	5.2	0.14	3.0	2.6	4.0							
3	2.4	1.99	2.6	2.2	26.7	0.63	3.2	3.0	7.8	0.072	2.7	2.5	5.5	0.15	3.2	2.6	6.7							
4	2.0	1.60	2.8	2.0	26.7	0.59	2.8	2.7	4.6	0.179	2.5	2.3	5.0	0.09	2.7	2.5	3.9							
5	2.5	3.23	2.6	2.2	43.9	0.85	3.0	2.8	10.2	0.095	2.6	2.4	5.4	0.20	2.9	2.4	8.3							
6	2.1	2.59	2.8	2.1	39.3	0.68	2.6	2.6	5.5	0.187	2.3	2.2	4.8	0.08	2.5	2.3	4.5							
7	4.5	0.06	3.5	3.0	7.9	0.06	5.0	4.7	4.7	0.001	4.2	4.0	5.2	0.24	5.2	4.5	5.8							
8	4.1	0.09	3.6	2.9	10.1	0.07	4.4	4.4	3.7	0.115	3.7	3.6	5.0	0.15	4.6	4.3	4.2							
9	5.4	2.98	3.8	3.2	27.4	1.15	5.0	4.8	8.5	0.087	4.6	4.4	5.3	0.28	4.9	4.3	6.5							
10	4.8	2.34	4.0	3.1	25.0	1.01	4.4	4.4	6.1	0.158	4.1	4.0	4.5	0.11	4.3	4.0	4.3							
11	5.2	4.23	3.7	3.2	40.0	1.37	4.7	4.4	9.4	0.073	4.3	4.1	5.3	0.36	4.5	3.9	7.0							
12	5.2	3.81	4.2	3.2	37.3	1.32	4.4	4.4	6.9	0.165	4.2	4.1	4.6	0.11	4.2	4.0	4.5							
13	44.9	-0.18	7.0	6.6	5.9	-0.30	9.5	9.3	5.7	-0.315	9.2	8.9	5.5	-0.36	9.4	8.8	6.2							
14	42.4	-0.25	6.8	6.7	4.6	-0.31	8.2	8.0	5.5	-0.341	7.8	7.7	5.2	-0.35	8.1	7.8	5.9							
15	39.1	1.82	7.2	6.5	11.0	0.86	9.9	9.5	7.1	-0.138	9.6	9.2	5.7	0.51	9.8	9.1	7.4							
16	36.6	1.42	7.0	6.6	8.8	1.05	8.5	8.4	7.2	-0.142	8.1	7.9	5.7	0.24	8.5	8.2	6.6							
17	38.1	3.19	7.2	6.5	15.2	1.49	9.9	9.6	7.7	-0.062	9.6	9.3	6.3	0.99	9.8	9.1	7.9							
18	36.9	2.55	7.0	6.5	12.0	1.54	8.5	8.4	8.2	-0.166	8.1	7.9	5.7	0.56	8.5	8.2	7.2							

Table 4

Bias, empirical standard deviation, average sandwich standard error and Type I error rate α under the null hypothesis ($\beta^* = 0$) when the exposure effects are monotonic and $P(Y_{00} = 1|X) = 0$.

	OLS					Cond. Monot.				Cond. Non-Monot.				Marginal				
	$E(Y)$ in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	α in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	α in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	α in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	α in %	
1	1.0	0.00	1.8	1.4	10.9	-0.04	2.2	2.1	5.8	-0.20	2.2	2.2	2.3	5.5	0.36	3.6	3.0	1.9
2	0.9	0.01	2.0	1.3	15.1	0.00	2.1	2.0	4.5	-0.15	2.1	2.1	1.9	4.9	0.19	2.8	2.5	1.0
3	1.6	1.44	2.2	1.7	26.8	0.05	2.4	2.4	5.3	-0.14	2.4	2.4	2.8	3.2	0.21	3.0	2.6	3.5
4	1.3	1.28	2.4	1.5	30.1	0.15	2.4	2.3	4.6	-0.01	2.4	2.4	2.2	4.7	0.13	2.4	2.3	1.7
5	1.7	2.34	2.2	1.8	40.2	0.12	2.3	2.2	5.6	-0.10	2.3	2.3	2.5	4.0	0.09	2.6	2.2	6.3
6	1.4	2.10	2.5	1.6	41.5	0.19	2.2	2.2	4.8	0.00	2.3	2.3	2.0	4.6	0.05	2.1	2.0	2.4
7	2.7	-0.01	3.0	2.2	10.9	-0.07	3.5	3.5	5.3	-0.26	3.5	3.5	4.2	3.2	0.31	4.8	4.4	1.6
8	2.3	0.02	3.1	2.1	13.9	0.00	3.4	3.3	4.8	-0.16	3.3	3.3	3.3	4.8	0.18	4.2	4.0	2.2
9	3.6	2.21	3.3	2.5	27.9	0.05	3.8	3.8	5.0	-0.18	3.8	3.8	4.5	2.7	0.22	4.4	4.1	2.1
10	3.0	1.94	3.6	2.3	30.5	0.09	3.7	3.7	4.7	-0.06	3.7	3.7	3.6	4.7	0.14	3.8	3.7	2.7
11	3.4	3.13	3.1	2.4	39.4	0.07	3.5	3.5	5.0	-0.13	3.5	3.5	4.1	3.1	0.12	3.8	3.5	2.6
12	3.4	3.18	3.8	2.4	40.7	0.11	3.8	3.8	4.6	-0.02	3.9	3.9	3.7	4.6	0.07	3.7	3.6	2.6
13	24.7	-0.13	6.5	4.6	11.8	-0.11	7.5	7.7	4.9	-0.11	7.5	7.5	8.4	3.9	-0.43	8.9	8.7	6.2
14	21.2	-0.17	6.4	4.6	11.8	-0.18	7.0	7.1	4.8	-0.18	7.0	7.0	7.1	4.9	-0.28	7.8	7.7	5.4
15	22.0	2.78	6.4	4.6	23.3	0.11	7.5	7.7	5.3	-0.04	7.5	7.5	8.4	4.2	0.53	9.0	8.8	7.4
16	18.7	2.35	6.4	4.5	21.5	-0.07	7.0	7.2	5.0	-0.13	7.0	7.0	7.1	4.9	0.34	7.9	7.9	6.1
17	20.7	4.11	6.3	4.6	30.2	0.17	7.5	7.7	5.5	0.00	7.5	7.5	8.3	4.4	0.81	9.0	8.7	7.1
18	19.1	3.62	6.4	4.5	28.4	-0.02	7.1	7.2	5.3	-0.10	7.1	7.1	7.2	5.4	0.56	8.0	7.9	6.4

Table 5

Bias, empirical standard deviation, average sandwich standard error and power under the alternative hypothesis ($\beta^* = 0.05$) when the exposure effects are monotonic and $P(Y_{00} = 1|X) = 0$.

OLS										Marginal					
Cond. Monot.					Cond. Non-Monot.										
$E(Y)$ in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	Power in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	Power in %	Power in %	Bias $\times 100$	E.SE $\times 100$	S.SE $\times 100$	Power in %	Power in %	Power in %
1	2.1	0.02	2.3	1.9	79.2	-0.09	2.9	2.9	52.2	-0.31	2.9	4.0	36.4	0.65	26.7
2	1.5	0.01	2.7	1.7	80.8	-0.09	3.0	2.9	52.7	-0.29	2.9	2.8	51.5	0.30	30.2
3	2.7	1.45	2.6	2.2	87.5	-0.06	3.2	3.2	46.4	-0.30	3.2	4.4	29.5	0.04	30.2
4	1.9	1.28	3.0	1.9	86.7	-0.03	3.2	3.2	45.1	-0.22	3.2	3.1	43.2	-0.09	32.6
5	2.8	2.36	2.6	2.2	92.6	-0.02	3.1	3.1	47.3	-0.30	3.1	4.1	30.8	-0.25	34.0
6	2.0	2.09	3.1	1.9	91.2	-0.01	3.2	3.1	45.9	-0.24	3.2	3.1	43.8	-0.33	35.1
7	3.8	0.01	3.3	2.6	60.8	-0.07	3.9	4.0	34.4	-0.31	3.9	5.7	20.2	0.45	17.8
8	3.0	0.02	3.6	2.3	63.5	-0.04	3.9	3.9	35.3	-0.23	3.9	3.8	34.2	0.22	21.6
9	4.7	2.23	3.5	2.8	77.2	0.01	4.2	4.3	31.9	-0.24	4.2	5.6	19.0	-0.01	19.3
10	4.0	3.11	4.1	2.6	82.5	0.07	4.4	4.4	30.3	-0.10	4.4	4.3	29.0	-0.19	23.1
11	4.5	3.16	3.4	2.7	85.8	0.03	4.0	4.1	34.4	-0.24	4.0	5.1	21.8	-0.25	22.1
12	3.3	2.78	3.8	2.4	84.7	0.05	4.0	4.0	33.2	-0.14	4.0	4.0	32.0	-0.26	25.2
13	25.7	-0.17	6.4	4.5	35.2	-0.19	7.3	7.5	16.2	-0.18	7.3	8.1	14.3	-0.52	15.4
14	21.8	-0.18	6.3	4.5	34.7	-0.19	6.8	6.9	18.0	-0.19	6.8	6.9	18.3	-0.28	16.8
15	23.1	2.75	6.3	4.5	51.7	0.07	7.5	7.6	17.0	-0.07	7.4	8.2	14.4	0.51	18.3
16	19.3	2.33	6.3	4.5	50.5	-0.11	7.0	7.1	17.6	-0.16	7.0	7.1	17.7	0.29	17.9
17	21.8	4.09	6.3	4.5	60.3	0.13	7.5	7.6	17.0	-0.03	7.4	8.1	14.2	0.77	18.2
18	19.7	3.61	6.3	4.4	59.2	-0.05	7.0	7.1	17.4	-0.12	7.0	7.1	17.5	0.54	18.8

Table 6

Bias, empirical standard deviation, average sandwich standard error and Type I error rate α under the null hypothesis ($\beta^* = 0$) when the exposure effects are non-monotonic.

	OLS										Cond. Monot.					Cond. Non-Monot.					Marginal Monot.					Marginal Non-Monot.				
	$E(Y)$ in %																													
	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %	Bias $\times 100$	E.S.E $\times 100$	S.S.E $\times 100$	α in %						
1	-0.06	3.4	3.3	5.5	4.68	4.9	4.8	23.8	-0.26	4.0	4.7	3.3	4.66	7.8	5.0	22.7	0.11	5.3	4.9	2.3										
2	-0.09	3.6	3.2	7.3	4.63	4.2	4.2	27.6	-0.24	3.9	3.8	4.6	4.58	5.9	4.6	23.4	0.04	4.6	4.5	3.0										
3	4.03	3.9	3.6	32.9	5.96	5.2	5.2	28.8	-0.03	4.4	5.1	3.2	5.38	6.9	4.8	29.1	0.27	4.7	4.5	3.5										
4	3.44	4.2	3.4	30.3	6.13	4.8	4.6	35.3	-0.11	4.3	4.2	4.3	5.15	5.2	4.3	30.7	0.02	4.2	4.1	2.9										
5	5.59	5.6	5.8	23.6	30.92	8.4	8.4	98.3	-0.11	6.6	7.1	4.0	27.96	12.3	7.6	92.3	0.16	7.3	7.1	4.6										
6	4.98	5.9	5.7	22.8	31.84	7.4	7.3	99.9	-0.19	6.4	6.3	5.1	27.80	9.0	6.9	98.5	0.04	6.5	6.5	4.8										