



Published in final edited form as:

J Am Stat Assoc. 2008 December 1; 103(484): 1693–1704. doi:10.1198/016214508000001084.

Multiply robust inference for statistical interactions

Stijn Vansteelandt

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

Tyler J. VanderWeele

Department of Health Studies, University of Chicago 5841 South Maryland Avenue, MC 2007,
Chicago, IL 60637

James M. Robins

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

Abstract

A primary focus of an increasing number of scientific studies is to determine whether two exposures interact in the effect that they produce on an outcome of interest. Interaction is commonly assessed by fitting regression models in which the linear predictor includes the product between those exposures. When the main interest lies in the interaction, this approach is not entirely satisfactory because it is prone to (possibly severe) bias when the main exposure effects or the association between outcome and extraneous factors are misspecified. In this article, we therefore consider conditional mean models with identity or log link which postulate the statistical interaction in terms of a finite-dimensional parameter, but which are otherwise unspecified. We show that estimation of the interaction parameter is often not feasible in this model because it would require nonparametric estimation of auxiliary conditional expectations given high-dimensional variables. We thus consider ‘multiply robust estimation’ under a union model that assumes at least one of several working submodels holds. Our approach is novel in that it makes use of information on the joint distribution of the exposures conditional on the extraneous factors in making inferences about the interaction parameter of interest. In the special case of a randomized trial or a family-based genetic study in which the joint exposure distribution is known by design or by Mendelian inheritance, the resulting multiply robust procedure leads to asymptotically distribution-free tests of the null hypothesis of no interaction on an additive scale. We illustrate the methods via simulation and the analysis of a randomized follow-up study.

Keywords

Double robustness; Gene-environment interaction; Gene-gene interaction; Longitudinal data; Semiparametric inference

1 Introduction

A primary focus of an increasing number of scientific studies is to determine whether two given exposures interact to produce their effect, i.e. to determine whether the effect of one exposure is modified by the second. For instance, in many longitudinal studies, the question of whether the time evolution of the response differs for subjects with different baseline characteristics/interventions is of primary interest. In genetic association studies of complex disorders, the discovery of gene-environment and gene-gene interactions is of great interest, because most complex disorders are thought to be caused by numerous genes and environmental factors, a subset of which may act synergistically. The development of robust

and powerful tests of gene-gene interaction is of special interest to geneticists, since, when the effect of one locus is modified by alleles at another locus, the power to detect a phenotypic association with the first locus can be greatly reduced unless the interaction is explicitly modelled (Cordell, 2002).

When the outcome is continuous or positive-constrained and uncensored, the presence of effect modification between exposures A_1 and A_2 is commonly assessed by fitting a linear or loglinear conditional mean model for the outcome Y , in which the linear predictor includes the product between these exposures. To be specific, let \mathbf{X} be a vector of measured pre-exposure variables such that conditioning on \mathbf{X} suffices to control for confounding when estimating the effects of A_1 and A_2 on outcome Y . Then the term β^* in the conditional mean model

$$E(Y|\mathbf{A}, \mathbf{X}) = g\left(\gamma_0^* + \gamma_1^* A_1 + \gamma_2^* A_2 + \gamma_3^* \mathbf{X} + \beta^* A_1 A_2\right) \quad (1)$$

with $\mathbf{A} = (A_1, A_2)'$ and $g(x) = x$ or $g(x) = e^x$ known, encodes the degree to which exposure A_2 modifies the effect of A_1 on outcome (on the scale g), and vice versa. Specifically, the choice $\beta^* = 0$ expresses that the effect of exposure A_1 on outcome is the same (on the scale g), regardless of the other exposure A_2 . It thus encodes the absence of effect modification (on scale g). More generally, one may fit a conditional mean model of the form

$$E(Y|\mathbf{A}, \mathbf{X}) = g\{m(\mathbf{A}, \mathbf{X}; \beta^*, \gamma^*)\} \quad (2)$$

where

$$m(\mathbf{A}, \mathbf{X}; \beta, \gamma) = q_3(\mathbf{A}, \mathbf{X}; \beta) + q_2(\mathbf{X}, A_2; \gamma_2) + q_1(\mathbf{X}, A_1; \gamma_1) + h(\mathbf{X}; \gamma_0)$$

with g defined as before, with $q_3(\mathbf{A}, \mathbf{X}; \beta)$ a known function smooth in β and satisfying $q_3(\mathbf{A}, \mathbf{X}; \beta) = 0$ when $A_1 A_2 = 0$, with $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$ and $h(\mathbf{X}; \gamma_0)$ known

functions smooth in $\gamma = (\gamma_0', \gamma_1', \gamma_2')$ (and γ_0 , γ_1 and γ_2 variation independent parameters), satisfying $q_1(\mathbf{X}, 0; \gamma_1) = q_2(\mathbf{X}, 0; \gamma_2) = 0$, with $\beta^* \in R^p$ and $\gamma^* \in R^q$ unknown parameters and with the joint law of (\mathbf{A}, \mathbf{X}) unrestricted. In this model, the term $q_3(\mathbf{A}, \mathbf{X}; \beta)$ encodes the statistical interaction between exposures A_1 and A_2 (possibly as a function of \mathbf{X}). Without loss of generality, we can require $q_3(\mathbf{A}, \mathbf{X}; \beta)$ to satisfy $q_3(\mathbf{A}, \mathbf{X}; 0) = 0$ so that $\beta^* = 0$ continues to encode the absence of statistical interaction. The functions $q_2(\mathbf{X}, A_2; \gamma_2)$ and $q_1(\mathbf{X}, A_1; \gamma_1)$ encode the main effects (possibly as functions of \mathbf{X}) of the exposures A_2 and A_1 , respectively. Finally, $h(\mathbf{X}; \gamma_0)$ encodes the main effect of the extraneous factors \mathbf{X} . For instance, model (1) is the special case in which $q_3(\mathbf{A}, \mathbf{X}; \beta) = \beta A_1 A_2$, $q_2(\mathbf{X}, A_2; \gamma_2) = \gamma_2 A_2$, $q_1(\mathbf{X}, A_1; \gamma_1) = \gamma_1 A_1$ and $h(\mathbf{X}; \gamma_0) = \gamma_0 + \gamma_3' \mathbf{X}$.

In observational studies, \mathbf{X} will typically be high-dimensional with a number of continuous components. For instance, in genetic association studies, \mathbf{X} might include a high-dimensional collection of substructure-informative loci (Epstein, Allen and Satten, 2007). This makes models for the main exposure effects $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$ and for the association $h(\mathbf{X}; \gamma_0)$ of extraneous factors \mathbf{X} with outcome prone to misspecification. These models are not in themselves of scientific interest when the primary goal is to test for statistical interaction between the exposures A_1 and A_2 . As such, standard tests of $\beta^* = 0$ and inference for statistical interaction under the above model is less than satisfactory. This is so because standard tests for statistical interaction tend to be heavily sensitive to

misspecification of the models for the main exposure effects (Greenland, 1993). In particular, they may fail to attain the nominal significance level when these nuisance models are misspecified. Similarly as demonstrated in the simulation study of Section 5, estimates of statistical interaction may be severely biased under misspecification of these nuisance models. As a consequence, in longitudinal studies, standard tests to determine whether the change in outcome mean over time depends on a particular baseline exposure may be compromised when the time evolution or main exposure effect is misspecified, or when important interactions with extraneous variables \mathbf{X} have been neglected or misspecified. In genetic association studies, tests for gene-gene or gene-environment interaction may be biased when the main effect of the gene/environment is incorrectly modelled (e.g. when a dominant genetic model was assumed but not appropriate) or interactions with extraneous confounders have been inadvertently omitted.

Our concern about the consequence of misspecifying the main exposure effects or the association between outcome and extraneous factors in statistical interaction tests is additionally motivated by a problem arising in the Sufficient-component cause framework (Rothman, 1976). It is well known that whether two variables statistically interact may depend on the particular model being used (e.g. on the chosen scale g in model (1)) (Mantel et al., 1977, Greenland, 1993). Specifically, two variables that have an interaction under one statistical model, may not have an interaction under a different model (e.g. with a different link function). When the outcome and exposures are dichotomous, it has been argued (Rothman, 1976; Koopman, 1981) that there is a natural, scale-independent way in which to assess the presence of interactions between two exposures, based on the Sufficient-component cause framework. This framework makes reference to the actual causal mechanisms involved in bringing about the outcome: when two or more binary causes participate in the same causal mechanism, it becomes proper to speak of Sufficient cause interactions. In recent work, VanderWeele and Robins (2007, 2008) derived various conditions which necessarily entail the presence of Sufficient cause interactions. When the exposure effects are assumed to be monotone (Greenland, 1993; VanderWeele and Robins, 2007), these conditions involve testing for effect modification on the risk difference scale. This scale suggests a Bernoulli regression model with linear link as the natural choice to test for sufficient cause interactions. Our interest in semiparametric tests now stems from the fact that such models are likely misspecified because for dichotomous outcomes they may not yield expected outcomes between 0 and 1. See the supplemental material (Vansteelandt et al., 2008) for further discussion of the relation between the estimators we derive below and interactions in the Sufficient cause framework (see also VanderWeele (2008)).

A number of approaches have been developed which avoid modelling the effects of extraneous factors and/or main exposure effects. Robins, Mark and Newey (1992) propose G-estimation which avoids modelling the effects of extraneous factors when a model is specified for the conditional mean exposure given these extraneous factors. Correlated data methods, such as conditional likelihood estimation (Verbeke, Spiessens and Lesaffre, 2001), regression of changes (Louis, 1988) and regression on within- and between-cluster effects (Neuhaus and Kalbfleisch, 1998) can be viewed as variants of this approach in the case of the linear link (Goetgeluk and Vansteelandt, 2007). However, all these approaches require a correct model for the main exposure effects.

In the context of longitudinal studies, Zeger and Diggle (1994) avoid modelling a main exposure effect (i.e. the time effect) via a backfitting algorithm which iterates between kernel estimation of the main time effect and generalized least squares estimation of the remaining parameters. When measurements are collected at discrete time points, Lin and Ying (2001) avoid nonparametric smoothing via a weighted least squares approach that is equivalent to G-estimation. Fan and Li (2004) relax these authors' restriction of

measurements being taken at discrete time points via an approximate regression of changes and profile least squares estimation. These approaches were specifically designed for longitudinal data and, with exception of the approximate regression of changes (Fan and Li, 2004), they require modelling the effects of extraneous factors \mathbf{X} . Multivariate adaptive regression splines (MARS) (Friedman, 1991) and interaction spline models (Chen, 1994) avoid parametric modelling assumptions on all exposures and extraneous factors. Although well suited for high-dimensional problems, they still suffer from the curse of dimensionality when the predictor space is large.

In this article, we develop a novel semiparametric approach that can perform well in moderate sized samples even when \mathbf{X} is high dimensional. In contrast to previous approaches, the performance guarantees offered by our new approach depend on the extent of prior knowledge concerning the joint exposure distribution $f(A_1, A_2|\mathbf{X})$ conditional on the covariates. Previous approaches do not make use of information concerning this law; as described below, by incorporating such information our estimators have certain multiple robustness properties described in detail in the next section. By incorporating information on the joint conditional exposure distribution $f(A_1, A_2|X)$, the class of estimators derived below also essentially encompasses a ‘propensity score’ approach to the estimation of interaction parameters.

Specifically suppose first that the joint law of A_1 and A_2 is known, as could be the case in either a clinical trial with A_1 and A_2 both randomly assigned or in a family-based gene-gene interaction study where the law of the genetic markers A_1 and A_2 is determined by Mendelian inheritance. In this setting, when g is the identity link, our approach delivers consistent and asymptotically normal (CAN) estimators of the interaction β^* and an asymptotically distribution free (ADF) test of the hypothesis $\beta^* = 0$ of no interaction, even when the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$ are all misspecified. In contrast, we prove that, even with $f(A_1, A_2|\mathbf{X})$ known, if the vector \mathbf{X} has a continuously distributed component and none of the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$ is guaranteed to be correct, it is impossible to obtain either a CAN estimator of β^* or an ADF test of no interaction when g is the exponential function.

Suppose next A_1 and A_2 are correctly assumed to be conditionally independent given \mathbf{X} so $f(A_1, A_2|\mathbf{X}) = f(A_1|\mathbf{X})f(A_2|\mathbf{X})$, but neither the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$ nor models $f(A_1|\mathbf{X}; \alpha_1)$ and $f(A_2|\mathbf{X}; \alpha_2)$ (with α_1 and α_2 variation independent parameters) for $f(A_1|\mathbf{X})$ and $f(A_2|\mathbf{X})$ are guaranteed correct. This would often be the case in population-based gene-environment interaction studies of a genetic marker score A_1 and an environmental exposure A_2 or in a population-based gene-gene interaction study where the two genetic markers A_1 and A_2 are unlinked, provided Sufficient information on ethnicity and geographic origin, or on parental genetic markers are recorded in \mathbf{X} to remove the effects of population stratification. In this setting, with g the identity link, we construct a CAN estimator of β^* under a union model that assumes at least one of the following four statements is true: (i) the models $f(A_1|\mathbf{X}; \alpha_1)$ and $f(A_2|\mathbf{X}; \alpha_2)$ are both correct, (ii) the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$ are all correct, (iii) the models $f(A_1|\mathbf{X}; \alpha_1)$ and $q_1(\mathbf{X}, A_1; \gamma_1)$ are both correct, or (iv) the models $f(A_2|\mathbf{X}; \alpha_2)$ and $q_2(\mathbf{X}, A_2; \gamma_2)$ are both correct. We refer to our estimation approach as quadruply robust as only one of (i)–(iv) need to hold to obtain a CAN estimator of β^* . For g the exponential link, it is only triply robust, delivering CAN estimators of β^* if at least one of (ii), (iii), or (iv) holds.

Finally suppose A_1 and A_2 are not known to be conditionally independent, given \mathbf{X} , and we therefore specify a model $f(A_1, A_2|\mathbf{X}; \alpha) = f(A_1|A_2, \mathbf{X}; \alpha_1) \times f(A_2|\mathbf{X}; \alpha_2)$ that allows for conditional dependence. None of the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$, $f(A_1|A_2, \mathbf{X}; \alpha_1)$, $f(A_2|\mathbf{X}; \alpha_2)$ is guaranteed correct. Then we shall see that, even for the identity link,

quadruply robust estimators are not possible, because there do not exist compatible models $f(A_1|A_2, \mathbf{X}; \alpha_1)$ and $f(A_2|A_1, \mathbf{X}; \alpha_2)$ with variation independent parameters. Specifically in this setting, with g the identity link, we construct a CAN estimator of β^* which we refer to as triply robust because it is CAN under a union model that assumes at least one of the following three statements is true: (i) the models $f(A_1|A_2, \mathbf{X}; \alpha_1)$ and $f(A_2|\mathbf{X}; \alpha_2)$ are both correct, (ii) the models $q_2(\mathbf{X}, A_2; \gamma_2)$, $q_1(\mathbf{X}, A_1; \gamma_1)$, $h(\mathbf{X}; \gamma_0)$ are all correct, (iii) the models $f(A_1|A_2, \mathbf{X}; \alpha_1)$ and $q_1(\mathbf{X}, A_1; \gamma_1)$ are both correct. For g the exponential link, our approach is also triply robust when (i) just above is replaced by the more restrictive condition that the models $f(A_1|A_2, \mathbf{X}; \alpha_1)$, $f(A_2|\mathbf{X}; \alpha_2)$ and $q_2(\mathbf{X}, A_1; \gamma_2)$ are all correct.

In summary, because of the multiple robustness property enjoyed by our approach, we would recommend that, when \mathbf{X} is high dimensional, it be used quite generally because an inference concerning an interaction effect under our approach, unlike under previous approaches, has multiple chances, rather than only one chance, to be correct or nearly correct.

The paper is organized as follows. In Section 2 we introduce semiparametric statistical interaction models. These parameterize the statistical interaction between exposures A_1 and A_2 (on the chosen scale g) as a function of exposures and extraneous variables \mathbf{X} in terms of a finite number of parameters, but leave the observed data law otherwise unrestricted. In particular, the proposed models leave the main effects of both exposures on the outcome unspecified, along with their interactions with extraneous variables. We examine properties of these models. We show that, due to the curse of dimensionality, no general ADF test for statistical interaction exists that is guaranteed to perform well in realistic-sized samples because estimation of the interaction parameters requires the auxiliary estimation of conditional expectations given high-dimensional variables. We therefore introduce parametric models that we characterize as ‘working’ models because they are not guaranteed to be correct. In Sections 3 and 4, we show how to construct the multiply robust estimators described above. In Section 3, we do so under the assumption that A_1 and A_2 are conditionally independent given \mathbf{X} . In Section 4 we allow for conditional dependence. We illustrate the performance of our methods via simulation studies in Section 5 and the analysis of a randomized follow-up study in Section 6.

2 Model and inference

Consider a study whose design calls for measurements on a vector of variables $(Y_i, \mathbf{A}_i, \mathbf{X}_i)$ to be recorded for each of $i = 1, \dots, n$ independent subjects. Here, Y_i is the outcome of interest, $\mathbf{A}_i = (A_{i1}, A_{i2})'$ is a vector of exposure variables A_{i1} and A_{i2} , and \mathbf{X}_i is a vector of extraneous variables, such as confounders for the association between exposure \mathbf{A}_i and outcome Y_i . The goal of the study is to assess whether the association between the exposure A_1 and the outcome Y is modified by A_2 on either an additive or multiplicative scale.

To investigate whether there exist ADF tests of the null hypothesis that the interaction parameter $\beta^* = \mathbf{0}$, we consider the semiparametric interaction model \mathcal{A} which relaxes some of the parametric restrictions of model (2). Specifically, model \mathcal{A} is defined by the conditional mean model

$$E(Y|\mathbf{A}, \mathbf{X}) = g\{m(\mathbf{A}, \mathbf{X}; \beta^*)\} \quad (3)$$

where

$$m(\mathbf{A}, \mathbf{X}; \beta) = q_3(\mathbf{A}, \mathbf{X}; \beta) + q_2(\mathbf{X}, A_2) + q_1(\mathbf{X}, A_1) + h(\mathbf{X})$$

with $q_3(\mathbf{A}, \mathbf{X}; \boldsymbol{\beta})$ defined as before, $q_2(\mathbf{X}, A_2)$, $q_1(\mathbf{X}, A_1)$ and $h(\mathbf{X})$ being unknown functions satisfying $q_1(\mathbf{X}, 0) = q_2(\mathbf{X}, 0) = 0$, with the joint law of (\mathbf{A}, \mathbf{X}) unrestricted, with $g(\cdot)$ known and either the identity or exponential function, and with $\boldsymbol{\beta}^* \in R^p$ an unknown parameter vector. For instance, we may postulate that

$$E(Y|\mathbf{A}, \mathbf{X}) = g\{\boldsymbol{\beta}^* A_1 A_2 + q_2(\mathbf{X}, A_2) + q_1(\mathbf{X}, A_1) + h(\mathbf{X})\} \quad (4)$$

for unknown functions $q_2(\mathbf{X}, A_2)$, $q_1(\mathbf{X}, A_1)$ and $h(\mathbf{X})$.

Theorem 1 gives the influence functions of regular asymptotically linear (RAL) estimators of $\boldsymbol{\beta}^*$ in model \mathcal{A} and will form the basis of our argument as to why estimation of $\boldsymbol{\beta}^*$ in model \mathcal{A} is infeasible when \mathbf{X} is high dimensional. The proof of this and other results are given in the supplemental material (Vansteelandt et al., 2008).

Theorem 1. If $\widehat{\boldsymbol{\beta}}$ is a regular asymptotically linear (RAL) estimator of $\boldsymbol{\beta}^*$ in model \mathcal{A} , then there exists a $p \times 1$ function $\mathbf{d}(\mathbf{A}, \mathbf{X})$ in the set \mathbf{D} of all $p \times 1$ functions of (\mathbf{A}, \mathbf{X}) satisfying

$$E\{\mathbf{d}(\mathbf{A}, \mathbf{X})|A_1, \mathbf{X}\} = E\{\mathbf{d}(\mathbf{A}, \mathbf{X})|A_2, \mathbf{X}\} = 0, \quad (5)$$

such that $\widehat{\boldsymbol{\beta}}$ has influence function $\mathbf{d}(\mathbf{A}, \mathbf{X})\epsilon(\boldsymbol{\beta})$, where $\epsilon(\boldsymbol{\beta}) = Y - m(\mathbf{A}, \mathbf{X}; \boldsymbol{\beta})$ when $g(x) = x$ and $\epsilon(\boldsymbol{\beta}) = Y \exp\{-m(\mathbf{A}, \mathbf{X}; \boldsymbol{\beta})\} - 1$ when $g(x) = e^x$. That is,

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = n^{-1/2} \sum_{i=1}^n \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \epsilon_i(\boldsymbol{\beta}) + o_p(1).$$

By standard results from semiparametric theory in Bickel et al. (1993), Theorem 1 implies that all regular and asymptotically linear (RAL) estimators of $\boldsymbol{\beta}^*$ in model \mathcal{A} can be obtained (up to asymptotic equivalence) as the solution $\widetilde{\boldsymbol{\beta}}(\mathbf{d})$ to the equation

$$\sum_{i=1}^n \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \epsilon_i(\boldsymbol{\beta}) = 0, \quad (6)$$

for some $\mathbf{d} \in \mathbf{D}$. The solution $\widetilde{\boldsymbol{\beta}}(\mathbf{d})$ to this equation is an infeasible estimator as the set of functions \mathbf{D} satisfying (5) depends on the unknown conditional law $f(\mathbf{A}_i|\mathbf{X}_i)$ of exposure \mathbf{A}_i , given \mathbf{X}_i , and $\epsilon_i(\boldsymbol{\beta})$ depends on the unknown functions $q_2(\mathbf{X}_i, A_{i2})$, $q_1(\mathbf{X}_i, A_{i1})$ and $h(\mathbf{X}_i)$. A feasible RAL estimator is not possible unless some of the unknowns $q_2(\mathbf{X}_i, A_{i2})$, $q_1(\mathbf{X}_i, A_{i1})$, $h(\mathbf{X}_i)$, and $f(\mathbf{A}_i|\mathbf{X}_i)$ can be consistently estimated. While smoothing methods could in principle be used, with the sample sizes found in practice, the data available to estimate the density $f(\mathbf{A}_i|\mathbf{X}_i)$ and the main effects $q_2(\mathbf{X}_i, A_{i2})$, $q_1(\mathbf{X}_i, A_{i1})$ and $h(\mathbf{X}_i)$ will be sparse when \mathbf{X}_i is a vector with at least several continuous components. As a consequence any feasible estimator of $\boldsymbol{\beta}^*$ under model \mathcal{A} will exhibit poor finite sample performance when the predictor space is large. It follows that in general inference about $\boldsymbol{\beta}^*$ in model \mathcal{A} is infeasible due to the curse of dimensionality and that dimension-reducing (e.g. parametric) working models must be used to estimate the unknowns $q_2(\mathbf{X}_i, A_{i2})$, $q_1(\mathbf{X}_i, A_{i1})$, $h(\mathbf{X}_i)$ and $f(\mathbf{A}_i|\mathbf{X}_i)$. In the following 2 sections, we demonstrate that multiply robust estimators of $\boldsymbol{\beta}^*$ are obtained when the parameters of these models are estimated in an appropriate fashion. In Section 3, we assume that A_1 and A_2 are conditionally independent given \mathbf{X} . This assumption is dropped in Section 4.

3 Conditionally independent exposures

As discussed in the introduction, there are important settings in which A_1 and A_2 are known to be conditionally independent given \mathbf{X} . Therefore define \mathcal{A}_{cip} like model \mathcal{A} , but with the additional assumption that $A_1 \perp\!\!\!\perp A_2 \mid \mathbf{X}$. Under this model, the set of estimating equations (6) with $\mathbf{d} \in \mathbf{D}$ can equivalently be rewritten as

$$0 = \sum_{i=1}^n [\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid A_{i1}, \mathbf{X}_i\} - E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid A_{i2}, \mathbf{X}_i\} + E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid \mathbf{X}_i\}] \epsilon_i(\beta) \tag{7}$$

where $\mathbf{d} = \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)$ is a member of the set of $p \times 1$ functions of $(\mathbf{A}_i, \mathbf{X}_i)$. The solution $\tilde{\beta}(\mathbf{d})$ to this equation is still an infeasible estimator for the reasons discussed previously.

We consider 4 possible dimension-reducing strategies based on working models. The first strategy is to postulate the parametric model (2), i.e., to postulate a parametric model \mathcal{M}_y for $q_2(\mathbf{X}, A_2) = q_2(\mathbf{X}, A_2; \gamma_2^*)$, $q_1(\mathbf{X}, A_1) = q_1(\mathbf{X}, A_1; \gamma_1^*)$ and $h(\mathbf{X}) = h(\mathbf{X}; \gamma_0^*)$ with $\gamma^* \equiv (\gamma_0^*, \gamma_1^*, \gamma_2^*)$ unknown finite dimensional parameters, and with γ_0, γ_1 and γ_2 variation independent. The second strategy is to postulate a parametric model \mathcal{M}_a for the conditional densities of A_j , given \mathbf{X} for $j = 1, 2$, i.e.

$$f(A_j \mid \mathbf{X}) = f(A_j \mid \mathbf{X}; \alpha_j^*),$$

where $f(A_1 \mid \mathbf{X}; \alpha_1)$ and $f(A_2 \mid \mathbf{X}; \alpha_2)$ are known densities smooth in variation independent parameters α_1 and α_2 , α_1^*, α_2^* are unknown finite-dimensional parameters and $\alpha^* \equiv (\alpha_1^*, \alpha_2^*)$. The third (fourth) strategy is to postulate the model $\mathcal{M}_{y_a j}, j = 1 (j = 2)$ that assumes $q_j(\mathbf{X}, A_j) = q_j(\mathbf{X}, A_j; \gamma_j^*)$ and $f(A_j \mid \mathbf{X}) = f(A_j \mid \mathbf{X}; \alpha_j^*)$.

Since we cannot be certain that any of these 4 models are correct, we aim to find an estimator $\widehat{\beta}$ of β^* that is guaranteed to be CAN when any one of them (but not necessarily more than 1 of them) is correct. That is, we wish to find estimators $\widehat{\beta}$ that are CAN in the union submodel $\mathcal{B}_{\text{cip}}^{\text{id}} \equiv \mathcal{A}_{\text{cip}}(\mathcal{M}_y \cup \mathcal{M}_a \cup \mathcal{M}_{y_a 1} \cup \mathcal{M}_{y_a 2})$ of model \mathcal{A}_{cip} that assumes that at least one of $\mathcal{M}_y, \mathcal{M}_a, \mathcal{M}_{y_a 1}$ and $\mathcal{M}_{y_a 2}$ is true. In line with Robins, Rotnitzky and van der Laan (2000), Robins and Rotnitzky (2001) and van der Laan and Robins (2003), we will refer to such estimators as quadruply robust and, more generally, as multiply robust estimators (Vansteelandt, Rotnitzky and Robins, 2007). Part (i) of Theorem 2 below shows that, under mild regularity conditions, when $g(\cdot)$ is the identity link, the estimators $\widehat{\beta}_{\text{cip}} \equiv \widehat{\beta}_{\text{cip}}(\mathbf{d})$ are multiply robust (in the sense of being CAN for β^* under model $\mathcal{B}_{\text{cip}}^{\text{id}}$) for $\widehat{\beta}_{\text{cip}}(\mathbf{d})$ the solution to

$$0 = \sum_{i=1}^n [\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid A_{i1}, \mathbf{X}_i; \widehat{\alpha}_2\} - E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid A_{i2}, \mathbf{X}_i; \widehat{\alpha}_1\} + E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) \mid \mathbf{X}_i; \widehat{\alpha}\}] \epsilon_i(\beta, \widehat{\gamma}(\beta)) \tag{8}$$

with $\epsilon_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = Y_i - m(\mathbf{A}_i, \mathbf{X}_i; \boldsymbol{\beta}, \boldsymbol{\gamma})$, $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)$ an arbitrary $p \times 1$ function of $(\mathbf{A}_i, \mathbf{X}_i)$, $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}'_1, \widehat{\alpha}'_2)'$, with $\widehat{\alpha}_j$ satisfying

$$0 = \sum_{i=1}^n \mathbf{H}_{ij}(\widehat{\alpha}_j) \equiv \sum_{i=1}^n \frac{\partial}{\partial \alpha_j} \ln f(A_{ij} | \mathbf{X}_i; \alpha_j) \Big|_{\alpha_j = \widehat{\alpha}_j}$$

for $j = 1, 2$, and $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})$ solving the system of equations

$$0 = \sum_{i=1}^n \mathbf{G}_{i0}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \equiv \sum_{i=1}^n \mathbf{c}_0(\mathbf{A}_i, \mathbf{X}_i) \epsilon_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{9}$$

$$0 = \sum_{i=1}^n \mathbf{G}_{i1}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \widehat{\alpha}_1) \equiv \sum_{i=1}^n [\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i) | A_{i2}, \mathbf{X}_i; \widehat{\alpha}_1\}] \epsilon_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{10}$$

$$0 = \sum_{i=1}^n \mathbf{G}_{i2}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \widehat{\alpha}_2) \equiv \sum_{i=1}^n [\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i) | A_{i1}, \mathbf{X}_i; \widehat{\alpha}_2\}] \epsilon_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{11}$$

for arbitrary vector functions $\mathbf{c}_0(\mathbf{A}_i, \mathbf{X}_i)$, $\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i)$ and $\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i)$ of the dimension of $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, respectively. The arguments of Robins and Rotnitzky (2001) imply that a necessary condition for the existence of such quadruply robust estimator of $\boldsymbol{\beta}^*$ in model

$A_{\text{cip}} \cap (\mathcal{M}_y \cup \mathcal{M}_a \cup \mathcal{M}_{ya1} \cup \mathcal{M}_{ya2})$ is that there exists an unbiased estimating equation for $\boldsymbol{\beta}^*$ (with non-trivial power against local alternatives) were any of the following four statements to hold: (1) $q_2(\mathbf{X}, A_2)$, $q_1(\mathbf{X}, A_1)$ and $h(\mathbf{X})$ are all known, (2) $f(A_2 | \mathbf{X})$ and $f(A_1 | \mathbf{X})$ are both known, (3) $q_1(\mathbf{X}, A_1)$ and $f(A_1 | \mathbf{X})$ are both known, (4) $q_2(\mathbf{X}, A_2)$ and $f(A_2 | \mathbf{X})$ are both known. The main step in the proof of Theorem 2 is showing that, for $j = 1, 2, 3, 4$, (8) is an unbiased estimating equation for $\boldsymbol{\beta}^*$ when statement j holds and the known values of the functions specified in statement j are substituted for their estimated values in (8). The proof is then completed by showing that all of the following are true: $f(A_{ij} | \mathbf{X}_i; \widehat{\alpha}_j)$ is a CAN estimator of $f(A_{ij} | \mathbf{X}_i)$ in models \mathcal{M}_a and \mathcal{M}_{ya_j} , $j = 1, 2$, $q_j(\mathbf{X}, A_j; \widehat{\boldsymbol{\gamma}}_j(\boldsymbol{\beta}^*))$ is a CAN estimator of $q_j(\mathbf{X}, A_j)$ in models \mathcal{M}_y and \mathcal{M}_{ya_j} , $j = 1, 2$, and $h(\mathbf{X}; \widehat{\boldsymbol{\gamma}}_0(\boldsymbol{\beta}^*))$ is a CAN estimator of $h(\mathbf{X})$ in model \mathcal{M}_y .

Part (i) of Theorem 2 further shows that when $g(\cdot)$ is the exponential link, all estimators $\widehat{\boldsymbol{\beta}}_{\text{cip}} \equiv \widehat{\boldsymbol{\beta}}_{\text{cip}}(\mathbf{d})$ obtained by solving (8) with $\epsilon_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = Y_i \exp\{-m(\mathbf{A}_i, \mathbf{X}_i; \boldsymbol{\beta}, \boldsymbol{\gamma})\} - 1$ are multiply robust in the sense of being CAN in the union model

$\mathcal{B}_{\text{cip}}^{\text{exp}} = A_{\text{cip}} \cap (\mathcal{M}_y \cup \mathcal{M}_{ya1} \cup \mathcal{M}_{ya2})$ when the above conditions hold. As discussed in the introduction, unlike under the identity link, the estimators $\widehat{\boldsymbol{\beta}}_{\text{cip}}(\mathbf{d})$ are not CAN in model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_a$. In fact, as mentioned above, a necessary condition for any estimator to be CAN in model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_a$, and thus in model $A_{\text{cip}} \cap (\mathcal{M}_y \cup \mathcal{M}_a \cup \mathcal{M}_{ya1} \cup \mathcal{M}_{ya2})$, is that an unbiased estimating equation for $\boldsymbol{\beta}^*$ (with non-trivial power against local alternatives) exists when $f(A_2 | \mathbf{X})$ and $f(A_1 | \mathbf{X})$ are known. But in Lemmas 1-3 of the supplemental materials

(Vansteelandt et al., 2008) we show that no such unbiased estimating equation need exist when g is the exponential function and \mathbf{X} has continuous components. The lack of an unbiased estimating equation in this setting is connected with the following non-collapsibility property of multiplicative interactions.

Remark

Non-Collapsibility of Multiplicative Interactions: Consider again model (4) and suppose \mathbf{A} and \mathbf{X} are independent. The model $E(Y|\mathbf{A}) = g(\kappa^* A_1 A_2 + \eta_1^* A_1 + \eta_2^* A_2 + \eta_0^*)$ is derived from model (4) by collapsing over \mathbf{X} . Note that this model is saturated when A_1 and A_2 are dichotomous. If $g(x) = x$, then $\beta^* = \kappa^*$, so additive interactions are collapsible over \mathbf{X} . However, a trivial calculation shows that if $g(x) = \exp(x)$, even $\beta^* = 0$ fails to imply $\kappa^* = 0$.

Theorem 2

Suppose that the regularity conditions stated in the supplemental materials (Vansteelandt et al., 2008) hold and that $\beta, \gamma, \alpha, \alpha_1$ and α_2 are variation independent.

(i) Then, when $g(x) = x$ ($g(x) = \exp(x)$), $\sqrt{n}(\widehat{\beta}_{\text{cip}} - \beta^*)$ is RAL under model $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ with influence function

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

and thus converges in distribution to a $N(0, \gamma)$, where

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

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

$$\begin{aligned} \mathbf{U}_i^*(\beta, \gamma, \alpha) &= \mathbf{U}_i(\beta, \gamma, \alpha) - E \left\{ \frac{\partial}{\partial \gamma} \mathbf{U}_i(\beta, \gamma, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \gamma} \mathbf{G}_i(\beta, \gamma, \alpha) \right\} \mathbf{G}_i(\beta, \gamma, \alpha) \\ &\quad - \left[E \left\{ \frac{\partial}{\partial \alpha} \mathbf{U}_i(\beta, \gamma, \alpha) \right\} - E \left\{ \frac{\partial}{\partial \gamma} \mathbf{U}_i(\beta, \gamma, \alpha) \right\} E^{-1} \left\{ \frac{\partial}{\partial \gamma} \mathbf{G}_i(\beta, \gamma, \alpha) \right\} \right] \\ &\quad \times E \left[\frac{\partial}{\partial \alpha} \mathbf{G}_i(\beta, \gamma, \alpha) \right] E^{-1} \left\{ \frac{\partial}{\partial \alpha} \mathbf{H}_i(\alpha) \right\} \mathbf{H}_i(\alpha) \end{aligned} \tag{12}$$

with $\mathbf{H}_i(\alpha) \equiv (\mathbf{H}'_{i1}(\alpha), \mathbf{H}'_{i2})'$ and $\mathbf{G}_i(\beta, \gamma, \alpha) \equiv (\mathbf{G}'_{i0}(\beta, \gamma), \mathbf{G}'_{i1}(\beta, \gamma, \alpha_1), \mathbf{G}'_{i2}(\beta, \gamma, \alpha_2))'$.

(ii) Furthermore, let $\widehat{\beta}(\mathbf{d}, \mathbf{G}_{(1)}, \mathbf{H}_{(1)})$ and $\widehat{\beta}(\mathbf{d}, \mathbf{G}_{(2)}, \mathbf{H}_{(2)})$ be 2 estimators of β^* under model $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ corresponding to the same index functions \mathbf{d} , but different unbiased estimating functions $\mathbf{G}_{(1)}$ and $\mathbf{G}_{(2)}$ for γ under model \mathcal{M}_γ and $\mathbf{H}_{(1)}$ and $\mathbf{H}_{(2)}$ for α^* under model \mathcal{M}_α . Then, $\sqrt{n} \{ \widehat{\beta}(\mathbf{G}_{(1)}, \mathbf{H}_{(1)}) - \widehat{\beta}(\mathbf{G}_{(2)}, \mathbf{H}_{(2)}) \} = o_p(1)$ at the intersection submodel $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_\gamma \cap \mathcal{M}_\alpha$.

Part (i) of Theorem 2 suggests that multiply robust estimators of β^* in model $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ can be obtained by solving an equation of the form (8). General results on doubly robust

estimation in Robins and Rotnitzky (2001) further imply that any regular CAN estimator of β^* in model $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ has the same asymptotic distribution as $\widehat{\beta}_{\text{cip}}(\mathbf{d})$ can be obtained in this way. Part (ii) of Theorem 2 suggests that the choice of estimators for α^* and γ^* has no impact on the efficiency of $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ when the models \mathcal{M}_y and \mathcal{M}_a are correctly specified. Thus the fact that γ_1^* and γ_2^* are estimated by G-estimators solving (10) and (11), respectively, rather than by their more efficient maximum likelihood estimators under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y$ has no effect on the asymptotic variance of $\widehat{\beta}_{\text{cip}}$ when the law of the data lies in $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y \cap \mathcal{M}_a$. Nonetheless, the use of such G-estimators is critical to control bias. Indeed, while the solution to (8) is a CAN estimator under model $\mathcal{A}_{\text{cip}} \cap (\mathcal{M}_y \cap \mathcal{M}_a)$ with γ^* is replaced by an arbitrary CAN estimator under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y$, it is not CAN under the less restrictive model $\mathcal{B}_{\text{cip}}^{\text{id}}$ (or $\mathcal{B}_{\text{cip}}^{\text{exp}}$).

It follows as a corollary of Theorem 4 in Section 4 that, when the residual outcome variance is constant in \mathbf{A} , i.e. $\text{Var}\{\epsilon(\beta^*|\mathbf{A}, \mathbf{X})\} = \sigma^2(\mathbf{X})$ for some function $\sigma^2(\mathbf{X})$, the efficient estimating equation at β in model \mathcal{A}_{cip} is obtained by replacing $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)$ in equation (8) with

$$\sigma^{-2}(\mathbf{X}_i) \frac{\partial}{\partial \beta} q_3(\mathbf{A}_i, \mathbf{X}_i; \beta)$$

For example, when $q_3(\mathbf{A}_i, \mathbf{X}_i; \beta) = q_3(\mathbf{X}_i; \beta)A_{i1}A_{i2}$, we obtain

$$0 = \sum_{i=1}^n \frac{\partial q_3(\mathbf{X}_i; \beta)}{\partial \beta} \{A_{i1} - E(A_{i1}|\mathbf{X}_i; \widehat{\alpha}_1)\} \{A_{i2} - E(A_{i2}|\mathbf{X}_i; \widehat{\alpha}_2)\} \frac{\epsilon_i(\beta, \widehat{\gamma}(\beta))}{\sigma^2(\mathbf{X}_i)} \tag{13}$$

It can be deduced from Robins and Rotnitzky (2001) that the semiparametric variance bound in models \mathcal{A}_{cip} and $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ are identical whenever the model $\mathcal{M}_y \cap \mathcal{M}_a$ is true, and thus that solving (13) then yields a semiparametric efficient estimator under model $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$ at the intersection model $\mathcal{M}_y \cap \mathcal{M}_a$. Note that (13) merely requires specifying the conditional means of A_1 and A_2 , given \mathbf{X} , and not the entire conditional distribution. In practice, unless the variance function $\sigma^2(\mathbf{X})$ is further assumed not to depend on \mathbf{X} , the unknown function $\sigma^2(\mathbf{X})$ in (13) must be replaced by an estimator.

The homoscedasticity assumption that $\mathcal{M}_y \cap \mathcal{M}_a$ does not depend on \mathbf{A} , may often be implausible and is logically impossible for count data with $g(x) = \exp(x)$. When this assumption fails, the efficient estimating equation at β^* in model \mathcal{A}_{cip} can be obtained following the methods developed in the next section.

4 Conditionally dependent exposures

In this section, we relax the previous assumptions by allowing for the exposures A_1 and A_2 to be conditionally dependent given \mathbf{X} .

4.1 Estimation

We first consider the special case of binary exposures. When A_1 and A_2 are dichotomous, as when testing for gene-gene interaction between 2 possibly linked bi-allelic markers each

with dominant or recessive mode of inheritance, then an arbitrary function $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)$ can be written as $A_{i2}A_{i1}\mathbf{d}_{11}(\mathbf{X}_i) + A_{i2}(1 - A_{i1})\mathbf{d}_{01}(\mathbf{X}_i) + (1 - A_{i2})A_{i1}\mathbf{d}_{10}(\mathbf{X}_i) + (1 - A_{i2})(1 - A_{i1})\mathbf{d}_{00}(\mathbf{X}_i)$ for given functions $\mathbf{d}_{kl}(\mathbf{X}_i)$, $k, l = 0, 1$. It follows that the set \mathbf{D} of functions (\mathbf{A}, \mathbf{X}) satisfying (5) is the set $\mathbf{D} = \{\mathbf{d}^\dagger(\mathbf{X})\Delta(\mathbf{A}, \mathbf{X}); \mathbf{d}^\dagger(\mathbf{X}) \in R^p\}$ where

$$\Delta(\mathbf{A}, \mathbf{X}) \equiv \frac{\frac{A_1A_2}{E\{A_1A_2|\mathbf{X}\}} + \frac{(1-A_1)(1-A_2)}{E\{(1-A_1)(1-A_2)|\mathbf{X}\}}}{-\frac{A_1(1-A_2)}{E\{A_1(1-A_2)|\mathbf{X}\}} - \frac{(1-A_1)A_2}{E\{(1-A_1)A_2|\mathbf{X}\}}} \tag{14}$$

$$= \{f(A|X)\}^{-1} [I\{A_1=A_2\} - I\{A_1 \neq A_2\}] \tag{15}$$

Hence the estimating equations (6) with $\mathbf{d} \in \mathbf{D}$ can equivalently be written as

$$0 = \sum_{i=1}^n \mathbf{d}^\dagger(\mathbf{X}_i) \Delta(\mathbf{A}_i, \mathbf{X}_i) \epsilon_i(\beta) \tag{16}$$

where \mathbf{d}^\dagger is a member of the set of all $p \times 1$ functions of \mathbf{X} .

When the exposures A_{i1} and A_{i2} are not both dichotomous and dependent conditional on \mathbf{X}_i , we can use the following characterization of the set \mathbf{D} of functions (\mathbf{A}, \mathbf{X}) satisfying (5).

Lemma (Tchetgen and Robins, 2008)—Let $f^*(\mathbf{A}|\mathbf{X}) = f^*(A_1|\mathbf{X})f^*(A_2|\mathbf{X})$ be any fixed density for $\mathbf{A}|\mathbf{X}$ with A_1 and A_2 conditionally independent given \mathbf{X} that is absolutely continuous with respect to the true density $f(\mathbf{A}|\mathbf{X})$. Then the set of functions \mathbf{D} satisfying (5) is the set $\{\mathbf{d}^\dagger(\mathbf{A}, \mathbf{X}, \mathbf{r}); \mathbf{r} = \mathbf{r}(\mathbf{A}, \mathbf{X}) \in R^p\}$ where

$$\mathbf{d}^\dagger(\mathbf{A}, \mathbf{X}, \mathbf{r}) = \frac{\frac{f^*(\mathbf{A}|\mathbf{X})}{f(\mathbf{A}|\mathbf{X})} [\mathbf{r}(\mathbf{A}, \mathbf{X}) - E^*\{\mathbf{r}(\mathbf{A}, \mathbf{X})|A_1, \mathbf{X}\} - E^*\{\mathbf{r}(\mathbf{A}, \mathbf{X})|A_2, \mathbf{X}\} + E^*\{\mathbf{r}(\mathbf{A}, \mathbf{X})|\mathbf{X}\}]}{\tag{17}}$$

and where the expectations $E^*(\cdot)$ are taken w.r.t. $f^*(\mathbf{A}|\mathbf{X})$.

When A_1 and A_2 are dichotomous and we choose $f^*(\mathbf{A}|\mathbf{X}) \equiv 1/4$ w.p.1 and $\mathbf{r}(\mathbf{A}, \mathbf{X}) = 4\mathbf{d}^\dagger(\mathbf{X}_i) [I\{A_1 = A_2\} - I\{A_1 \neq A_2\}]$, we obtain $\mathbf{d}^\dagger(\mathbf{A}, \mathbf{X}, \mathbf{r}) = \mathbf{d}^\dagger(\mathbf{X}_i)\Delta(\mathbf{A}_i, \mathbf{X}_i)$, thus establishing equation (16) as a special case of equation (17). For non-dichotomous A_1 and A_2 , given any (user-supplied) density $f^*(\mathbf{A}|\mathbf{X}) = f^*(A_1|\mathbf{X})f^*(A_2|\mathbf{X})$ satisfying the conditions of the lemma, we can apply equation (17) to an arbitrary (user-supplied) function, say $\mathbf{d}^{(1)}(\mathbf{A}, \mathbf{X})$, to obtain a function $\mathbf{d}^\dagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ that satisfies equation (5).

An alternative way to map an arbitrary function $\mathbf{d}^{(1)}(\mathbf{A}, \mathbf{X})$ to an element of \mathbf{D} is to apply the alternating conditional expectations (ACE) algorithm (Breiman and Friedman, 1985; Bickel et al., 1993). This is an iterative algorithm which, starting from $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i)$, computes the repeating conditional expectations

$$\mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}^{(2m-1)}(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{d}^{(2m-1)}(\mathbf{A}_i, \mathbf{X}_i) | A_{i1}, \mathbf{X}_i\} \tag{18}$$

$$\mathbf{d}^{(2m+1)}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) - E\{\mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) | A_{i2}, \mathbf{X}_i\} \tag{19}$$

for $m = 1, 2, \dots$ until convergence at $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)}) = \lim_{m \rightarrow \infty} \mathbf{d}^{(2m+1)}(\mathbf{A}_i, \mathbf{X}_i)$. The function $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ then satisfies equation (5). Although both in \mathbf{D} , $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ and $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ will generally differ. The function $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ exists in closed form and is easy to compute. In contrast, the function $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ cannot be expressed in closed form when A_1 and A_2 are both continuous, unlike when A_1 and/or A_2 is discrete (Bickel et al., 1993); even so, in general, $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ remains more difficult to compute than $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$. Furthermore, we shall see below that (a weighted version) of $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$ is needed to obtain a locally semiparametric efficient estimator.

Unlike in the previous section, there do not exist compatible models $f(A_1|A_2, \mathbf{X}; \alpha_1)$ and $f(A_2|A_1, \mathbf{X}; \alpha_2)$ with variation independent parameters when conditional dependence between both exposures is allowed for. Inference for β^* therefore cannot be made robust to misspecification of either one of these conditional densities, and thus no consistent estimators can be obtained under model $\mathcal{A} \cap (\mathcal{M}_{y_{a1}} \cup \mathcal{M}_{y_{a2}} \cup \mathcal{M}_{y_{a1}})$.

Remark—More precisely, it can be shown that we could construct compatible models for $f(A_1|A_2, \mathbf{X})$ and $f(A_2|A_1, \mathbf{X})$ with variation independent parameters α_1 and α_2 if we assume that, for chosen values a_{10}, a_{20} , the generalized odds ratio function $\rho(A_1, A_2, \mathbf{X}) = f(A_1|A_2, \mathbf{X})f(A_1 = \alpha_{10}|A_2 = \alpha_{20}, \mathbf{X}) / \{f(A_1 = \alpha_{10}|A_2, \mathbf{X})f(A_1|A_2 = \alpha_{20}, \mathbf{X})\}$ is a known function, simply by specifying models $f(A_1|A_2 = \alpha_{20}, \mathbf{X}; \alpha_1)$ and $f(A_2|A_1 = \alpha_{10}, \mathbf{X}; \alpha_2)$ for $f(A_1|A_2 = \alpha_{20}, \mathbf{X})$ and $f(A_2|A_1 = \alpha_{10}, \mathbf{X})$. However, in practice, the assumption that $\rho(A_1, A_2, \mathbf{X})$ is known would never be reasonable, except in the special case that the generalized odds ratio function is the constant function 1, which is equivalent to again assuming A_1 and A_2 are conditionally independent given \mathbf{X} . However, if we did not restrict attention to variation independent models, it is possible to drop the assumption that $\rho(A_1, A_2, \mathbf{X})$ is known and specify a model $\rho(A_1, A_2, \mathbf{X}; \zeta)$ for $\rho(A_1, A_2, \mathbf{X})$ depending on a parameter vector ζ . Then the model $\rho(A_1, A_2, \mathbf{X}; \zeta)$ together with the aforementioned models $f(A_1|A_2 = \alpha_{20}, \mathbf{X}; \alpha_1)$ and $f(A_2|A_1 = \alpha_{10}, \mathbf{X}; \alpha_2)$ induce compatible models for $f(A_1|A_2, \mathbf{X})$ and $f(A_2|A_1, \mathbf{X})$ with the parameter ζ occurring in both. We could then construct consistent estimators of β^* when either the model for $f(A_1|A_2, \mathbf{X})$ or the model for $f(A_2|A_1, \mathbf{X})$ is correct, because, using methods described in Chen (2007) and Tchetgen and Robins (2008), the common parameter ζ can be consistently estimated if either the model $f(A_1|A_2 = \alpha_{20}, \mathbf{X}; \alpha_1)$ or the model $f(A_2|A_1 = \alpha_{10}, \mathbf{X}; \alpha_2)$ is correct.

We will therefore conduct inference for β^* under model $\mathcal{B}^{\text{id}} \equiv \mathcal{A}(\mathcal{M}_y \cup \mathcal{M}_a \cup \mathcal{M}_{y_{a1}})$, where we redefine \mathcal{M}_a to be a parametric model for the conditional density of \mathbf{A} , given \mathbf{X} , of the form

$$f(\mathbf{A}|\mathbf{X}) = f(\mathbf{A}|\mathbf{X}; \alpha^*) = f(A_1|\mathbf{X}, A_2; \alpha_1^*) f(A_2|\mathbf{X}; \alpha_2^*).$$

Here, $f(A_1|\mathbf{X}, A_2; \alpha_1)$ and $f(A_2|\mathbf{X}; \alpha_2)$ are known densities smooth in α_1 and α_2 , and

$\alpha^* = (\alpha_1^*, \alpha_2^*)'$ is an unknown finite-dimensional parameter. Further, we define

$\mathcal{M}_{y_{a2}}^* \equiv \mathcal{M}_{y_{a2}} \cap \mathcal{M}_a$. Let $\tilde{\alpha} = (\tilde{\alpha}_1', \tilde{\alpha}_2')'$ satisfy

$$0 = \sum_{i=1}^n \mathbf{H}_{i1}(\widehat{\alpha}_1) \equiv \sum_{i=1}^n \frac{\partial}{\partial \alpha_1} \ln f(A_{i1}|A_{i2}, \mathbf{X}_i; \alpha_1)_{|\alpha_1 = \widehat{\alpha}_1} \quad (20)$$

$$0 = \sum_{i=1}^n \mathbf{H}_{i2}(\widehat{\alpha}_2) \equiv \sum_{i=1}^n \frac{\partial}{\partial \alpha_2} \ln f(A_{i2}|\mathbf{X}_i; \alpha_2)_{|\alpha_2 = \widehat{\alpha}_2} \quad (21)$$

Hence $\widehat{\alpha}_2$ is the MLE of α_2^* under model \mathcal{M}_a , while $\widehat{\alpha}_1$ is the MLE of α_1^* under both models \mathcal{M}_a and $\mathcal{M}_{y_{a1}}$.

Let $\Delta(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$, $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)}; \widehat{\alpha})$ and $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)}; \widehat{\alpha})$ be $\Delta(\mathbf{A}_i, \mathbf{X}_i)$, $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$, and $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)})$, except with the expectations now evaluated under $f(\mathbf{A}, \mathbf{X}; \widehat{\alpha})$. Given $\mathbf{d}^\ddagger(\mathbf{X}_i)$ and $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i)$, let $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ be $\mathbf{d}^\ddagger(\mathbf{X}_i) \Delta(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ when A_{i1} and A_{i2} are binary and let $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ be either $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)}; \widehat{\alpha})$ or $\mathbf{d}^\ddagger(\mathbf{A}_i, \mathbf{X}_i, \mathbf{d}^{(1)}; \widehat{\alpha})$ otherwise, where the dependence of $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ on \mathbf{d}^\ddagger or $\mathbf{d}^{(1)}$ is suppressed. In all cases, the function $\mathbf{d}(\widehat{\alpha}) = \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ is an element of $\mathbf{D}(\widehat{\alpha})$, where the set $\mathbf{D}(\widehat{\alpha})$ is defined like the set \mathbf{D} but with $f(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ replacing $f(\mathbf{A}_i|\mathbf{X}_i)$ in equation (5).

Theorem 3 below shows that when $g(\cdot)$ is the identity link, the estimators $\widehat{\beta} \equiv \widehat{\beta}(\mathbf{d}(\widehat{\alpha}))$ for a given $\mathbf{d}(\widehat{\alpha}) = \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha})$ are multiply robust (in the sense of being CAN for β^* under model \mathcal{B}^{id}), where $\widehat{\beta}(\mathbf{d}(\widehat{\alpha}))$ solves

$$0 = \sum_{i=1}^n \mathbf{U}_i(\beta, \widehat{\gamma}(\beta), \widehat{\alpha}) = \sum_{i=1}^n \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha}) \epsilon_i(\beta, \widehat{\gamma}(\beta)) \quad (22)$$

with $\widehat{\beta} \equiv \widehat{\beta}(\mathbf{d}(\widehat{\alpha}))$ still defined as in Section 3. Theorem 3 further shows that when $g(\cdot)$ is the exponential link, the estimators $\widehat{\beta} \equiv \widehat{\beta}(\mathbf{d}(\widehat{\alpha}))$ obtained by solving (22) with $\epsilon_i(\beta, \gamma) = Y_i \exp\{-m(\mathbf{A}_i, \mathbf{X}_i; \beta, \gamma)\} - 1$ are multiply robust in the sense of being CAN in the union model \mathcal{B}^{exp} .

Theorem 3—Suppose that the regularity conditions stated in the supplemental materials (Vansteelandt et al., 2008) hold and that, $\beta_1, \gamma_2, \gamma_1$ and α_2 are variation independent. Suppose $\mathbf{d}(\widehat{\alpha}) \in \mathbf{D}(\widehat{\alpha})$. Then Parts 1 and 2 of Theorem 2 continue to hold with $\widehat{\beta} \equiv \widehat{\beta}(\mathbf{d}(\widehat{\alpha}))$ replacing $\widehat{\beta}_{\text{cip}}$ and model $\mathcal{B}^{\text{id}}(\mathcal{B}^{\text{exp}})$ replacing $\mathcal{B}_{\text{cip}}^{\text{id}}(\mathcal{B}_{\text{cip}}^{\text{exp}})$, with $A_{i1}(\widehat{\alpha}_1)$ and $A_{i2}(\widehat{\alpha}_2)$ now defined as in (20) and (21).

We propose two practical strategies for implementing the ACE algorithm when when A_1 and A_2 are both continuous. The first strategy is a numerical integration approach whereby the integrals

$$E\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \alpha) | A_{ij}, \mathbf{X}_i; \alpha\} = \int \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \alpha) f(A_{ij}' | A_{ij}, \mathbf{X}_i; \alpha) dA_{ij}'$$

for $j, j' = 1, 2, j \neq j'$ in the ACE algorithm are approximated via numerical integration methods, such as the composite Simpson's rule (with α replaced by $\tilde{\alpha}$). This requires that we can evaluate the function $\mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i; \alpha)$ (and thus that we run the ACE algorithm) at a Sufficient number M of points $(a_{i11}, a_{i21}), \dots, (a_{i1M}, a_{i2M})$, spread across the support of (A_1, A_2) . These may be chosen for each given \mathbf{X}_i separately by drawing a random sample from the joint distribution of (A_{i1}, A_{i2}) , given \mathbf{X}_i , and should additionally include the observed data points at the given \mathbf{X}_i . Note that we opt for the composite Simpson's rule because this merely requires knowing the function values of $\mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i; \alpha)$ at the selected M points.

The second strategy is an ad-hoc approach which involves postulating separate high-dimensional models for the conditional expectations in (18) and (19) and fitting these each time using standard regression techniques (thus without postulating a model for $f(\mathbf{A}|\mathbf{X})$), as in Breiman and Friedman (1985). A drawback of this approach is that it does not guarantee congenial models for the conditional expectations in (18) and (19) (i.e. there may be no joint law $f(\mathbf{A}|\mathbf{X})$ for which the postulated conditional expectations (18) and (19) hold for $m = 1, 2, \dots$). Nevertheless, we recommend this approach for data analysis because the numerical integration approach is computer intensive, generally does not lead to improved results in simulation studies (see Section 5) and, to the best of our knowledge, its convergence properties have not been studied, unlike those of the ad-hoc approach (Breiman and Friedman, 1985). Furthermore, while there may be concerns over using automatic model fitting for the conditional expectations in (18) and (19) in the sense that these may be more likely misspecified, these concerns are mitigated to some extent by the robustness property of our estimators.

Remark—Using (12) to estimate the asymptotic variance of $\tilde{\beta}$ requires knowing the derivative $E\{\partial \mathbf{U}_i(\beta, \gamma, \alpha)/\partial \alpha\}$. This is difficult when $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \alpha)$ is obtained via the ACE algorithm because it then has no closed-form expression. However, a variance estimate can still be obtained by noting that, as shown in the supplemental materials (Vansteelandt et al., 2008), under models $B_{id} (B_{exp})$,

$$E\left\{\frac{\partial}{\partial \alpha} \mathbf{U}_i^*(\beta^*, \tilde{\gamma}(\beta^*), \alpha)\right\}_{|\alpha=\tilde{\alpha}} = -E\left\{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \tilde{\alpha}) \mathbf{S}_i(\tilde{\alpha}) \epsilon_i(\beta^*, \tilde{\gamma}(\beta^*))\right\} \tag{23}$$

where $\mathbf{S}_i(\tilde{\alpha})$ is the score for under model \mathcal{M}_α evaluated at $\tilde{\alpha}$.

Expression (23) is not useful under the ad-hoc implementation of the ACE algorithm because the score $\mathbf{S}_i(\alpha)$ is then unknown. In that case, one might for simplicity choose to ignore estimation of α^* when calculating the standard error of $\tilde{\beta}$. Indeed, Theorem 2.3 in van der Laan and Robins (2003) assures that, if model \mathcal{M}_α holds, ignoring efficient estimation of α^* leads to conservative inferences for β^* under model $B_{id} (B_{exp})$.

Furthermore, because $E\{\partial \mathbf{U}_i(\beta^*, \tilde{\gamma}, \alpha)/\partial \alpha\}_{|\alpha=\tilde{\alpha}}=0$ and $E\{\partial \mathbf{G}_i(\beta^*, \tilde{\gamma}, \alpha)/\partial \alpha\}_{|\alpha=\tilde{\alpha}}=0$ when model \mathcal{M}_γ is correctly specified, estimation of α^* does not affect the distribution of our estimator for β^* at model \mathcal{M}_γ (see expression (12)). This approach is not attractive, however, because simulation studies in Section 5 show that ignoring estimation of α^* in constructing our variance estimator may imply a serious loss of power when model \mathcal{M}_γ is misspecified. We therefore recommend the nonparametric bootstrap for inference under ad-hoc implementations of the ACE algorithm, as the bootstrap always provides a consistent estimator of the asymptotic variance under our assumptions.

4.2 Local semiparametric efficiency

We now consider how to obtain locally semiparametric efficient estimators. The key to doing so is the following characterization of the efficient score in model A. Let $\sigma^2(\mathbf{A}, \mathbf{X}) \equiv \text{Var} \{ \epsilon(\beta^*) | \mathbf{A}, \mathbf{X} \}$. In Theorem 4, we show that when A_{i1} and A_{i2} are binary, the efficient score for β^* is $\mathbf{S}_{\text{eff}} = \mathbf{d}_{\text{opt}}(\mathbf{A}_i, \mathbf{X}_i)_{\epsilon_i(\beta^*)}$ in model A with $\mathbf{d}_{\text{opt}}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}_{\text{opt}}^\dagger(\mathbf{X}_i) \Delta(\mathbf{A}_i, \mathbf{X}_i)$ and

$$\mathbf{d}_{\text{opt}}^\dagger(\mathbf{X}_i) = E \left\{ \Delta^2(\mathbf{A}_i, \mathbf{X}_i) \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i | \mathbf{X}_i) \right\}^{-1} E \left\{ \Delta(\mathbf{A}_i, \mathbf{X}_i) \frac{\partial}{\partial \beta} q_3(\mathbf{A}_i, \mathbf{X}_i; \beta)_{\beta=\beta^*} | \mathbf{X}_i \right\}.$$

When A_{1i} and A_{2i} are both continuous, the efficient score $\mathbf{S}_{\text{eff}} = \mathbf{d}_{\text{opt}}(\mathbf{A}_i, \mathbf{X}_i)_{\epsilon_i(\beta^*)}$ does not exist in closed form. However, regardless of the sample spaces of A_1 and A_2 , we show in Theorem 4 that $\mathbf{d}_{\text{opt}}(\mathbf{A}_i, \mathbf{X}_i) = \lim_{m \rightarrow \infty} \mathbf{d}^{(2m+1)}(\mathbf{A}_i, \mathbf{X}_i)$ is always the function to which the v -weighted ACE algorithm defined by

$$\mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}^{(2m-1)}(\mathbf{A}_i, \mathbf{X}_i) - v(\mathbf{A}_i, \mathbf{X}_i) \frac{E \{ \mathbf{d}^{(2m-1)}(\mathbf{A}_i, \mathbf{X}_i) | A_{i1}, \mathbf{X}_i \}}{E \{ v(\mathbf{A}_i, \mathbf{X}_i) | A_{i1}, \mathbf{X}_i \}} \tag{24}$$

$$\mathbf{d}^{(2m+1)}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) - v(\mathbf{A}_i, \mathbf{X}_i) \frac{E \{ \mathbf{d}^{(2m)}(\mathbf{A}_i, \mathbf{X}_i) | A_{i2}, \mathbf{X}_i \}}{E \{ v(\mathbf{A}_i, \mathbf{X}_i) | A_{i2}, \mathbf{X}_i \}} \tag{25}$$

for $m = 1, 2, \dots$, converges for the choices $v(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i)$ and $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i) \partial q_3(\mathbf{A}_i, \mathbf{X}_i; \beta)_{\beta=\beta^*} / \partial \beta$. The unweighted ACE algorithm defined by equations (18) and (19) is the special case of the v -weighted ACE algorithm with $v(\mathbf{A}_i, \mathbf{X}_i) = v^*(\mathbf{X}_i)$ only a function of \mathbf{X}_i . For any $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i) \in R^p$ and any always-positive function $v(\mathbf{A}_i, \mathbf{X}_i)$, the v -weighted algorithm, like the unweighted algorithm converges to a function $\mathbf{d}_v(\mathbf{A}_i, \mathbf{X}_i; \mathbf{d}^{(1)})$, that satisfies equation (5). This last statement follows from the following arguments. First,

$\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) - v(\mathbf{A}_i, \mathbf{X}_i) \frac{E \{ \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) | A_{ij}, \mathbf{X}_i \}}{E \{ v(\mathbf{A}_i, \mathbf{X}_i) | A_{ij}, \mathbf{X}_i \}}$ for $j \in \{1, 2\}$, is the orthogonal projection of the univariate function $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)$ on the closed linear subspace $\Lambda_j = \{ \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i); E \{ \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) | A_{ij}, \mathbf{X}_i \} = \mathbf{0} \}$ in the Hilbert space of functions of $(\mathbf{A}_i, \mathbf{X}_i)$ with inner product $\langle \mathbf{d}_1, \mathbf{d}_2 \rangle \equiv E [\{ v(\mathbf{A}_i, \mathbf{X}_i) \}^{-1} \mathbf{d}_1(\mathbf{A}_i, \mathbf{X}_i) \mathbf{d}_2(\mathbf{A}_i, \mathbf{X}_i)]$. It then follows from a theorem of Von Neumann (Bickel et al., 1993, p.436) that $\mathbf{d}_v(\mathbf{A}_i, \mathbf{X}_i; \mathbf{d}^{(1)})$ is the projection of $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i)$ on the linear space $\Lambda = \Lambda_1 \cap \Lambda_2$, which is precisely the subspace satisfying equation (5).

We now explain how to obtain a locally efficient estimator of β^* . Consider the model

$$\text{Var}(\epsilon(\beta^*) | \mathbf{A}, \mathbf{X}) = \sigma^2(\mathbf{A}, \mathbf{X}; \boldsymbol{\eta}^*) \tag{26}$$

where $\sigma^2(\mathbf{A}, \mathbf{X}; \boldsymbol{\eta})$ is a known function, smooth in $\boldsymbol{\eta}$, and $\boldsymbol{\eta}^*$ is an unknown parameter vector. Let $\widehat{\boldsymbol{\eta}}$ satisfy

$$0 = \sum_{i=1}^n \mathbf{H}_{i3}(\widehat{\boldsymbol{\eta}}) = \sum_{i=1}^n \mathbf{s}(\mathbf{A}_i, \mathbf{X}_i) \left\{ \epsilon_i^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\beta}})) - \sigma^2(\mathbf{A}_i, \mathbf{X}_i; \widehat{\boldsymbol{\eta}}) \right\}$$

where $\mathbf{s}(\mathbf{A}_i, \mathbf{X}_i)$ is a vector of user-supplied functions of the dimension of $\boldsymbol{\eta}$, and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\mathbf{d}(\widehat{\alpha}))$ for a given $\mathbf{d}(\widehat{\alpha}) = \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha}) \in \mathbf{D}(\widehat{\alpha})$. Note that, with any positive $v(\mathbf{A}_i, \mathbf{X}_i)$ and any $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i)$ as input, the v -weighted ACE algorithm, based on $f(\mathbf{A}_i | \mathbf{X}_i; \widehat{\alpha})$ rather than on $f(\mathbf{A}_i | \mathbf{X}_i)$, outputs a function $\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha}) \in \mathbf{D}(\widehat{\alpha})$.

Theorem 4—(i) The efficient score for β^* in model \mathcal{A} is $\mathbf{d}_{opt}(\mathbf{A}_i, \mathbf{X}_i) \in_i(\beta^*)$

1. with $\mathbf{d}_{opt}(\mathbf{A}_i, \mathbf{X}_i) = \mathbf{d}_{opt}^\dagger(\mathbf{X}_i) \Delta(\mathbf{A}_i, \mathbf{X}_i)$ for binary A_{i1} and A_{i2} ;
2. with $\mathbf{d}_{opt}(\mathbf{A}_i, \mathbf{X}_i) = \lim_{m \rightarrow \infty} \mathbf{d}^{(2m+1)}(\mathbf{A}_i, \mathbf{X}_i)$ in the v -weighted ACE algorithm with $v(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i)$ and $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i) \times \partial q_3(\mathbf{A}_i, \mathbf{X}_i; \beta) |_{\beta=\beta^*} / \partial \beta$ in general.

(ii) Let $\widehat{\boldsymbol{\beta}}(\mathbf{d}(\widehat{\alpha}))$ and $\widehat{\boldsymbol{\beta}}(\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}))$ solve (22) where $\mathbf{d}(\widehat{\alpha}) \equiv \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha}) \in \mathbf{D}(\widehat{\alpha})$ and $\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}) = \mathbf{d}_{opt}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\alpha}, \widehat{\eta}) \in \mathbf{D}(\widehat{\alpha})$ is the function to which the v -weighted ACE algorithm based on $f(\mathbf{A}_i | \mathbf{X}_i; \widehat{\alpha})$ converges for $v(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\eta})$ and $\mathbf{d}^{(1)}(\mathbf{A}_i, \mathbf{X}_i) = \sigma^{-2}(\mathbf{A}_i, \mathbf{X}_i; \widehat{\eta}) \partial q_3(\mathbf{A}_i, \mathbf{X}_i; \beta) |_{\beta=\beta^*} / \partial \beta$. Then, $\widehat{\boldsymbol{\beta}}(\mathbf{d}(\widehat{\alpha}))$ and $\widehat{\boldsymbol{\beta}}(\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}))$ are RAL estimators in models \mathcal{B}^{id} or \mathcal{B}^{exp} . If, in addition, the true distribution of the data lies in the intersection submodel $\mathcal{A} \cap \mathcal{M}_y \cap \mathcal{M}_\alpha$ and model (26) holds, then the difference between the asymptotic variance matrices of $\widehat{\boldsymbol{\beta}}(\mathbf{d}(\widehat{\alpha}))$ and $\widehat{\boldsymbol{\beta}}(\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}))$ is non-negative definite, with the asymptotic variance of $\widehat{\boldsymbol{\beta}}(\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}))$ equalling $\{\text{Var}(S_{eff})\}^{-1} = [\text{Var}\{\mathbf{d}_{opt}(\mathbf{A}_i, \mathbf{X}_i) \in_i(\beta^*)\}]^{-1}$.

It follows from Part (ii) of Theorem 4 that $\widehat{\boldsymbol{\beta}}(\mathbf{d}_{opt}(\widehat{\alpha}, \widehat{\eta}))$ is a locally semiparametric efficient of β^* in model A (and following the general results in Robins and Rotnitzky (2001) then also in models B^{id} or B^{exp}) at the intersection submodel in which model (26) and models \mathcal{M}_y and \mathcal{M}_α all hold.

5 Simulation study

We conducted a simulation experiment to evaluate the behaviour in finite samples of the multiply robust estimators for statistical interaction parameters. Each experiment was based on 1000 replications of random samples of size 500 generated as follows. Exposures were generated as $A_1 = 1 + X + \delta U + \epsilon_1$ and $A_2 = 1 - X + \delta U + \epsilon_2$, where X, U, ϵ_1 and ϵ_2 are four independent standard normal variates and where δ was set to 0 or 1 to represent settings without and with conditionally independent exposures, given X , respectively. The outcome was generated as $Y = -1 + A_1 + A_2 - A_1 A_2 + X + \lambda(A_1 - A_2)X + \epsilon$, where ϵ is a standard normal variate and λ was set to 0 and -2.

In each simulation experiment, 4 estimators were calculated under model A with $q_3(A, X; \beta) = \beta A_1 A_2$. The first is an ordinary least squares (OLS) estimate under working model \mathcal{M}_y , which is defined by $q_2(A_2, X; \gamma_2^*) = \gamma_2^* A_2$, $q_1(A_1, X; \gamma_1^*) = \gamma_1^* A_1$ and $h(X; \gamma_0^*) = \gamma_{0,0}^* + \gamma_{0,1}^* X$. The second is an efficient G-estimate (G) (Robins, Mark and Newey, 1992), assuming that $q_2(A_2, X; \gamma_2^*) = \gamma_2^* A_2$ and $q_1(A_1, X; \gamma_1^*) = \gamma_1^* A_1$ and that, in addition, model \mathcal{M}_y holds or model M_{2G} holds, which is defined by (correctly specified) second-order linear regression models for $E(A_j | X)$, $j = 1, 2$ and a (correctly specified) third-order linear regression model for $E(A_1 A_2 | X)$. The third (CI) is obtained by solving (7) assuming that either model \mathcal{M}_y holds, or model M_{2CI} holds, which is defined by second-order linear regression models for $E(A_j | X)$, $j = 1, 2$ and the assumption that $A_1 \perp\!\!\!\perp A_2 | X$. The fourth (ACE) is obtained by solving (22)

under working model M_y , having first applied the ACE algorithm under the ad-hoc strategy of Section 4, using linear regression models for the conditional expectations in (18) and (19) which involve third-order poly-nomials in A_j ($j = 1$ and 2 , respectively) along with interactions with X , and third-order polynomials in X , and assuming a constant residual variance.

The results of the simulation study are summarized in Table 1 and Figure 1. Variance estimates were obtained via the ordinary nonparametric bootstrap based on 500 resamples for the ACE- and CI-estimate, using sandwich estimators for the G-estimator and using the Fisher information matrix for the OLS estimator. Reported coverage for the ACE and CI-estimates is based on 95% basic bootstrap intervals.

The results indicate that the ad-hoc implementation of the ACE algorithm yields unbiased estimators for the statistical interaction parameter under each of the four data-generating models. This is because the chosen conditional mean models for (18) and (19) in the ACE algorithm were sufficiently flexible to yield approximately correctly specified models. None of the other estimators shares this property: the OLS and G-estimates are biased whenever the main effects of A_1 and A_2 are misspecified (i.e. $\lambda \neq 0$), although the OLS estimates are more severely biased. Estimate CI is biased when, in addition, A_1 and A_2 are conditionally dependent, given X (i.e. $\lambda \neq 0$ and $\delta \neq 0$). The price to pay for the increased robustness of our estimators is a loss of efficiency. This loss can be important when the conditional mean model for the outcome is correctly specified, but overall, reasonable efficiency was obtained with the semiparametric approach. Estimates obtained via the ACE algorithm were substantially more precise than those obtained under the conditional independence assumption (CI) when the conditional mean model M_y was incorrectly specified, even when the exposures were conditionally independent given X . This is in conformity with the fact that, whenever model M_y is incorrectly specified, one may gain efficiency by estimating the exposure distribution under a model that fails to impose a priori known restrictions such as the conditional independence of the exposures (van der Laan and Robins, 2003). Curiously, the CI estimate is much more precise than the estimate obtained using the ACE algorithm when the conditional independence assumption fails and the conditional mean model M_y is correctly specified. This is because the index function $d(A_i, X_i)$ of the CI estimate is much more variable than the corresponding function obtained via the ACE algorithm when the exposures are conditionally dependent, given X . For example, in the extreme case that $A_1 = A_2$ w.p.1, $d(A_i, X_i) = 0$ is the only solution to (5) and hence no multiply robust root- n estimators for β^* exist under laws at which $A_1 = A_2$ w.p.1, whilst the estimating functions in (7) yield root- n estimators of β^* under such laws (however, only when the conditional mean model M_y is correctly specified). We also evaluated doubly robust estimators obtained by replacing γ^* by an ordinary least squares estimate (instead of a G-estimate). This had no impact on the bias and variance of the doubly robust estimators obtained by the ACE algorithm because these are based on correctly specified models for the exposure distribution $f(A|X)$. However, it did impact the bias and variance of the CI estimators under the simulation experiments with conditional dependence: $(\delta, \lambda) = (1, 0)$ (bias $-3 \cdot 10^{-4}$, bootstrap variance 0.00034, empirical variance 0.00035, Type I error rate 0.063) and $(\delta, \lambda) = (1, -2)$ (bias 0.23, bootstrap variance 0.017, empirical variance 0.019, Type I error rate 0.48).

Table 1 further shows results for the numerical integration approach of Section 4 with $m = 100$ and using (correctly specified) third-order linear regression models with constant variance and normal errors for the conditional distributions of A_1 , given (A_2, X) , and of A_2 , given (A_1, X) . The complexity of these models warrants use of the bootstrap for inference. However, no bootstrap-based variance estimates are reported because the numerical

integration approach was extremely time-consuming. Table 1 shows that the obtained estimates (NI) are more efficient than those obtained under the ad-hoc strategy when the conditional mean model for the outcome is correctly specified, but they are biased and have greater imprecision otherwise. This is due to numerical approximation error and the fact that, when the conditional mean model for the outcome is incorrectly specified, the estimation procedure relies more heavily on restriction (5), and thus on the numerical integration. Indeed, the bias of the estimates diminished noticeably upon repeating the numerical integration approach for $m = 200$, at the expense of a serious increase in computation time.

6 Data analysis

To illustrate the methods, we re-analyze data from a placebo-controlled randomized trial conducted in 1989-1990 in the UK to study blood pressure reduction, as described in Goetghebeur and Lapp (1997). The trial started with a run-in period of 4 weeks whereby all patients received placebo tablets and after which they were randomized to 4 weeks of one of two active treatments (A or B) or placebo. Diastolic blood pressure measurements were taken every 2 weeks. For illustration, we analyze the subset of 105 patients randomized to treatment A or placebo, ignoring 2 patients with missing outcome data. Figure 2 shows a profile plot of the data.

Let Y denote diastolic blood pressure, A_1 be a binary variable taking the value 1 for patients randomized to the experimental treatment A during the active study period and 0 otherwise, A_2 denote time in days since enrollment into the study and X measure centered body weight (in kg). Fitting the following model

$$E(Y|A_1, A_2, X) = \gamma_0 + \gamma_1 X + \gamma_2 A_2 + \beta_1 A_1 A_2 + \beta_2 A_1 A_2 X$$

using generalized estimating equations with exchangeable working correlation, yields $\widehat{\beta}_1 = -0.12$ (SE 0.029) and $\widehat{\beta}_2 = 0.0089$ (SE 0.0044). This suggests that the average change in diastolic blood pressure per day is 0.12 (95% CI 0.067 - 0.18) higher in the experimental treatment arm than in the placebo arm among patients of average body weight. This difference reduces with 0.0089 (95% CI 0.00027 - 0.017) per kg increase in body weight.

To examine whether these results continue to hold, even under possible misspecification of the time evolution and possible interactions of time with body weight, we use the methods developed in this article. These methods, including the efficient score expressions, continue to hold for correlated data provided that scalar outcomes Y_i are replaced by vectors that contain all outcome measurements for the i th cluster, and likewise for the remaining data A_i , X_i , etc. In the analyses below, we let $\sigma^2(X_i; \beta, \gamma)$ in (26) be the working covariance matrix obtained via generalized estimating equations and use the bootstrap for inference. Under the valid assumption that the observation times are independent of assigned treatment, given body weight, we now estimate that the average change in diastolic blood pressure per day is 0.14 (95% CI 0.075 - 0.18) higher in the experimental treatment arm than in the placebo arm among patients with average body weight. This difference reduces with 0.014 (95% CI -0.0044 - 0.023) per kg increase in body weight. These estimates are distribution-free because $E(A_1|X) = E(A_1)$ by the fact that randomization happened independently of body weight, and likewise, because $E(A_2|X) = E(A_2)$ by the fact that the study design was completely balanced in time. In particular, the obtained estimates will be valid, even if the main effects of time and body weight (and possible interactions between both) have been incorrectly specified. Using the ACE algorithm we obtain similar, but slightly less efficient estimates of 0.13 (95% CI 0.075 - 0.18) for the main effect and 0.017 (95% CI -0.0078 -

0.028) for the interaction. All results confirm that the reduction in blood pressure over time is significantly different in both treatment arms. Given that standard estimates for statistical interactions can be very sensitive to the model for the main effects, these new results are more trustworthy, at the expense of a relatively limited degree of precision loss.

Alternatively we could have used randomization inference (Rosenbaum, 2002) for estimating the 2 considered interactions. This would also yield distribution-free inference by the fact that the ‘main’ effect of A_1 can be assumed to be zero and thus its misspecification is not at issue. Even so, the proposed multiply robust estimators enjoy a greater attraction than estimators obtained from a computationally more involved randomization inference, because they are available in closed-form. Furthermore, it is unclear how randomization inference could protect against misspecification of both main exposure effects. For further discussions on randomization inference versus semiparametric inference, see Robins (2002).

7 Conclusion

In this article, we have developed multiply robust estimators for statistical interaction parameters indexing additive or multiplicative conditional mean models. The estimators in the additive model are especially attractive in settings where the distribution of exposure given the extraneous covariates \mathbf{X} is known, as is generally the case in randomized follow-up studies and family-based genetic association studies, because they can be used to construct asymptotically distribution-free tests of the no-interaction hypothesis, even when the vector \mathbf{X} is high dimensional with continuous components. This makes our approach distinct from existing approaches, which ignore prior information on the exposure distribution. Our proposed approach can be used quite generally, even when, as in most observational studies, no such prior information is available, because an inference concerning an interaction effect under our approach has multiple chances, rather than only one chance, to be correct or nearly correct. In future work, we will apply the proposed estimators to develop scale-invariant interaction tests based on the Sufficient-component cause framework. In addition, we will extend the proposed methods to allow for ascertainment conditions, such as frequently encountered in genetic association studies, whereby data are sampled conditional on the outcome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Eric Tchetgen, the Editor, the Associate Editor and 2 referees for helpful comments. The first author acknowledges support from IAP research network grant nr. P06/03 from the Belgian government (Belgian Science Policy).

References

- Chamberlain G. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*. 1987; 34:305–334.
- Chen HY. A semiparametric odds ratio model for measuring association. *Biometrics*. 2007; 63:413–421. [PubMed: 17688494]
- Chen ZH. Fitting multivariate regression-functions by interaction spline models. *Journal of the Royal Statistical Society Series B*. 1993; 55:473–491.
- Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*. 2002; 11:2463–2468. [PubMed: 12351582]

- Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case-control studies. *American Journal of Human Genetics*. 2007; 80:921–930. [PubMed: 17436246]
- Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*. 2004; 99:710–723.
- Flanders WD. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *European Journal of Epidemiology*. 2006; 21:847–853. [PubMed: 17048084]
- Friedman J. Multivariate adaptive regression splines (with discussion). *Annals of Statistics*. 1991; 19:1–141.
- Goetghebeur E, Lapp K. The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Applied Statistics*. 1997; 46:351–364.
- Goetgeluk S, Vansteelandt S. Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics*. 2007 doi: 10.1111/j.1541-0420.2007.00944.x.
- Greenland S. Basic problems in interaction assessment. *Environmental Health Perspectives*. 1993; 101:59–66. [PubMed: 8206043]
- Greenland S, Brumback B. An overview of relations among causal modelling methods. *International Journal of Epidemiology*. 2002; 31:1030–1037. [PubMed: 12435780]
- Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work Environment and Health*. 1988; 14:125–129.
- Koopman JS. Interaction between discrete causes. *American Journal of Epidemiology*. 1981; 113:716–724. [PubMed: 7234861]
- Lin DY, Ying Z. Semiparametric and nonparametric analysis of longitudinal data (with discussion). *Journal of the American Statistical Association*. 2001; 96:103–126.
- Louis TA. General methods for analyzing repeated measures. *Statistics in Medicine*. 1988; 7:29–45. [PubMed: 3281207]
- Mantel N, Brown C, Byar DP. Tests for homogeneity of effect in an epidemiologic investigation. *American Journal of Epidemiology*. 1977; 106:125–129. [PubMed: 888814]
- Miettinen OS. Causal and preventive interdependence: Elementary principles. *Scandinavian Journal of Work Environment and Health*. 1982; 8:159–168.
- Miettinen, OS. *Modern Epidemiology*. John Wiley; New York: 1985.
- Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. 1998; 54:638–645.
- Robins JM, Mark SD, Newey WK. Estimating exposure effects by modeling the expectation of exposure conditional on confounders. *Biometrics*. 1992; 48:479–495. [PubMed: 1637973]
- Robins JM, Rotnitzky A, van der Laan M. Comment on ‘On Profile Likelihood’ by S. A. Murphy and A. W. van der Vaart. *Journal of the American Statistical Association*. 2000; 95:431–435.
- Robins JM, Rotnitzky A. Inference for semiparametric models: Some questions and an answer - Comments. *Statistica Sinica*. 2001; 11:920–936.
- Robins JM. Comment on ‘Covariance adjustment in randomized experiments and observational studies’, by P. R. Rosenbaum. *Statistical Science*. 2002; 17:286–327.
- Robins, J.; Li, L.; Tchetgen, E.; van der Vaart, A. *IMS Collections Probability and Statistics: Essays in Honor of David A. Freedman*. Vol. Vol. 2. 2008. Higher order influence functions and minimax estimation of nonlinear functionals; p. 335-421.
- Rosenbaum PR. Covariance adjustment in randomized experiments and observational studies. *Statistical Science*. 2002; 17:286–304.
- Rothman KJ. *Causes*. *American Journal of Epidemiology*. 1976; 104:587–592. [PubMed: 998606]
- Rothman, KJ.; Greenland, S. *Modern Epidemiology*. Lippincott-Raven; Philadelphia, PA: 1998.
- Tchetgen, ET.; Robins, JM. Technical report. Dept. of Epidemiology, Harvard school of Public Health; 2008. On doubly robust estimation in a semiparametric odds ratio model.
- Umbach DH, Weinberg CR. The use of case-parent triads to study joint effects of genotype and exposure. *American Journal of Human Genetics*. 2000; 66:251–261. [PubMed: 10631155]

- van der Laan, MJ.; Robins, JM. Unified Methods for Censored Longitudinal Data and Causality. Springer-Verlag; New-York: 2003.
- van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; Cambridge: 1998.
- VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology*. 2008 in press.
- VanderWeele TJ, Robins JM. The identification of synergism in the Sufficient-component cause framework. *Epidemiology*. 2007; 18:329–339. [PubMed: 17435441]
- VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for Sufficient cause interactions. *Biometrika*. 2008; 95:49–61.
- Vansteelandt S, Rotnitzky A, Robins JM. Estimation of regression models for the mean of repeated outcomes under non-ignorable non-monotone non-response. *Biometrika*. 2007; 94:841–860.
- Vansteelandt, S.; VanderWeele, TJ.; Robins, JM. Supplemental materials for “Multiply robust inference for statistical interactions”. 2008. http://www.amstat.org/publications/jasa/supplemental_materials
- Verbeke G, Spiessens B, Lesaffre E. Conditional linear mixed models. *The American Statistician*. 2001; 55:25–34.
- Zeger SL, Diggle PJ. Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*. 1994; 50:689–699. [PubMed: 7981395]

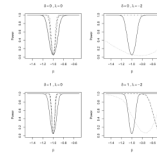


Figure 1. Power of 4 statistical interaction tests of the null hypothesis that $\beta^* = -1$: OLS (long-short dashed), G (long dashed), CI (dotted), ACE (solid).

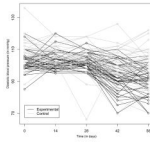


Figure 2.
Profile plot of diastolic blood pressure in 2 treatment arms.

Table 1

Bias, variance, empirical variance and Type I error rate (α) of tests performed at the 5% significance level.

(δ, λ)	Estimator	Bias	Variance	Empirical Var	α
(0,0)	ACE	$9 \cdot 10^{-5}$	0.0023	0.0023	0.045
	CI	$6 \cdot 10^{-4}$	0.0025	0.0024	0.037
	G	$8 \cdot 10^{-4}$	0.00067	0.00072	0.066
	OLS	$4 \cdot 10^{-4}$	0.00041	0.00042	0.058
	NI	$6 \cdot 10^{-4}$	-	0.0010	-
(0,-2)	ACE	$9 \cdot 10^{-5}$	0.0023	0.0023	0.045
	CI	-0.024	0.14	0.11	0.066
	G	1.32	0.0056	0.0058	1.00
	OLS	2.40	0.0050	0.019	1.00
	NI	0.063	-	0.028	-
(1,0)	ACE	$3 \cdot 10^{-4}$	0.0013	0.0013	0.050
	CI	$3 \cdot 10^{-3}$	0.00059	0.00052	0.053
	G	$1 \cdot 10^{-4}$	0.00028	0.00029	0.054
	OLS	$5 \cdot 10^{-5}$	0.00023	0.00023	0.050
	NI	$5 \cdot 10^{-4}$	-	0.0011	-
(1,-2)	ACE	0.00029	0.0013	0.0013	0.050
	CI	0.12	0.016	0.016	0.24
	G	0.57	0.0062	0.0063	1.00
	OLS	1.33	0.0056	0.029	1.00
	NI	-0.30	-	0.035	-