Harvard University
Harvard University Biostatistics Working Paper Series

Year 2012 *Paper* 142

Multiple-Robust Estimation of an Odds Ratio Interaction

Eric J. Tchetgen Tchetgen*

*Harvard University, etchetge@hsph.harvard.edu

Original Article

Multiple-Robust Estimation of an Odds Ratio Interaction

Eric J. Tchetgen Tchetgen

Departments of Epidemiology and Biostatistics,

Harvard University

Correspondence: Eric J. Tchetgen Tchetgen, Department of Epidemiology, Harvard School of Public Health 677 Huntington Avenue, Boston, MA 02115.

Abstract

An important scientific goal of genetic epidemiology studies is increasingly to determine whether an interaction between a genetic factor and an environmental factor or two genetic factors is present in the effect that they produce on the risk of a disease outcome. In such studies, interaction is commonly assessed by fitting a logistic regression model to case-control or cohort data, in which the linear predictors includes on the log-odds scale, the product between the two factors of interest. Unfortunately, inferences on an interaction using standard logistic regression methods are prone to bias due to model mis-specification of main exposure effects or of the association model between extraneous factors and the outcome. In this paper, an alternative semiparametric logistic regression model is considered, which postulates the statistical interaction in terms of a finite-dimensional parameter, but which is otherwise unspecified. We show that estimation is generally not feasible in this model because of the curse of dimensionality associated with the required estimation of auxiliary conditional densities given high-dimensional covariates. We thus consider 'multiply robust estimation' and propose a more general model which assumes at least one of several working models holds. We illustrate the methods via simulation and the analysis of an Israeli ovarian case-control study.

KEY WORDS: Double robustness; Gene-environment interaction; Gene-gene interaction;logistic regression; Semiparametric inference.

1 Introduction

A common scientific aim of genetic epidemiology studies is to determine whether a genetic variant and an environmental factor, or two genes interact in the effect that they produce on the risk of a disease outcome. When the outcome is binary, the presence of effect modification between exposures A_1 and A_2 is commonly assessed by fitting a logistic regression 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. In observational studies, $\mathbf{X}$ will typically be high-dimensional with a number of continuous components. Throughout this article, we will therefore consider $\mathbf{X}$ to be a high-dimensional vector. It then follows that the term β^* in the logistic model

$$\operatorname{logit}\Pr(Y=1|\mathbf{A},\mathbf{X}) = \gamma_0^* + \gamma_1^* A_1 + \gamma_2^* A_2 + \gamma_3^{*'} X + \beta^* A_1 A_2 \tag{1}$$

with $\mathbf{A} = (A_1, A_2)'$, encodes the degree to which exposure A_2 modifies the effect of A_1 on the odds-ratio scale of the outcome risk, and vice versa. Specifically, the choice $\beta^* = 0$ expresses that the effect of exposure A_1 on the outcome is the same on the odds ratio scale, regardless of the other exposure A_2. It thus encodes the absence of effect modification on the odds ratio scale. The odds ratio scale is an attractive scale to estimate associations in general, because it is invariant to alterations in the marginal distributions of $(Y, \mathbf{X})$ or of $(\mathbf{A}, \mathbf{X})$. As a consequence, it accommodates data collected under a variety of commonly employed epidemiological designs, in particular, simple random sample designs and case-control designs unmatched or matched on some or all the components of $\mathbf{X}$. To allow for generality in our exposition, suppose instead of (1) one fits a logistic model of the form

$$\text{logit} \Pr(Y = 1|\mathbf{A}, \mathbf{X}) = q_3 (\mathbf{A}, \mathbf{X}; \beta^*) + q_2 (\mathbf{X}, A_2; \gamma_2^*) + q_1 (\mathbf{X}, A_1; \gamma_1^*) + h(\mathbf{X}; \gamma_0^*) \tag{2}$$

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 $\gamma^* = (\gamma_0^{*\prime}, \gamma_1^{*\prime}, \gamma_2^{*\prime})'$ (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 $\gamma^* \in R^p$ 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 on the odds ratio scale, the statistical interaction between exposures A_1 and A_2 (possibly as a function of $\mathbf{X}$), in other words, $q_3 (\mathbf{A}, \mathbf{X}; \beta^*)$ is equal to :

$$\log \frac{\mathbf{OR}_{Y,A|X} (1, \mathbf{A}|\mathbf{X})}{\mathbf{OR}_{Y,A|X} (1, (A_1, A_2 = 0) |\mathbf{X})} = \log \frac{\mathbf{OR}_{Y,A|X} (1, \mathbf{A}|\mathbf{X})}{\mathbf{OR}_{Y,A|X} (1, (A_1 = 0, A_2) |\mathbf{X})}$$

where $\mathbf{OR}_{Y,A|X} (\mathbf{A}|\mathbf{X})$ is the conditional odds ratio association relating Y and $\mathbf{A}$ within levels 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 $h(\mathbf{X}; \gamma_0) = \gamma_0^* + \gamma_3^{*\prime}\mathbf{X}$.

In the following, models for the main exposure effects $q_2(\mathbf{X}, A_2; \gamma_2^*)$, $q_1(\mathbf{X}, A_1; \gamma_1^*)$ and for main effect $h(\mathbf{X}; \gamma_0^*)$ of extraneous factors $\mathbf{X}$ on the outcome are not of primary scientific interest and constitute auxiliary models. Our primary goal is to construct an estimator and/or test for a statistical interaction between the exposures A_1 and A_2. Standard tests of the null hypothesis of no interaction, i.e. $\beta^* = 0$, such as the fully parametric logistic regression approach described above, are not entirely satisfactory, because they require the analyst consistently estimates these auxiliary quantities. In fact, Tchetgen Tchetgen (2010) demonstrated in a simulation that standard logistic regression estimation of a statistical interaction can be severely biased when a main effect is misspecified. His finding agrees with related simulation results by Vansteelandt et al (2008) who considered the performance of standard regression analysis for evaluating an interaction on a linear or a log-linear scale. To remedy the serious limitation of standard regression analysis, Vansteelandt et al developed multiply robust estimators which they show have favorable theoretical properties, and which were also shown to outperform standard methods in simulation studies. For improved robustness, their approach uses a model for the conditional density of $\mathbf{A}$ given $\mathbf{X}$ in addition to a working model for the outcome regression of Y on $(\mathbf{A}, \mathbf{X})$. However, to remain valid, their approach only requires correct specification of some but not all of auxiliary models. In fact, they showed that for an additive interaction, their approach is consistent and asymptotically normal (CAN), provided that at least one of the following four conditions holds:

a) the outcome regression model of Y on $(\mathbf{A}, \mathbf{X})$ is correctly specified, or

b) a model for the effect of A_1 on the mean of Y given $(A_2, \mathbf{X})$,and the density of A_1 given $(A_2, \mathbf{X})$ are both correct, or

c) a model for the effect of A_2 on the mean of Y given $(A_1, \mathbf{X})$, and the density of A_2 given $(A_1, \mathbf{X})$ are both correct, or

d) the density of $\mathbf{A}$ given $\mathbf{X}$ is correct.

Thus, their estimator of an additive interaction is quadruply robust. In other words, unlike a standard regression analysis which offers a single opportunity for obtaining valid inferences by requiring that a) holds, the method of Vansteelandt et al (2008) gives four such opportunities. For making inferences about a multiplicative interaction, their approach which is triply robust remains CAN when at least one of a), b) or c) holds. Because an inference concerning an interaction effect under their approach, unlike under previous approaches, has multiple chances, rather than only one chance, to be correct or nearly correct, Vansteelandt et al (2008) recommended it be used quite generally, particularly when as in most observational studies, $\mathbf{X}$ is high dimensional. However, the methods developed in Vansteelandt et al (2008) do not apply if the outcome is dichotomous and the hypothesized interaction operates on the odds ratio scale.

In this article, we consider a semiparametric theory for odds ratio interactions. Specifically, we consider a semiparametric logistic regression model which postulates a statistical interaction in terms of a finite-dimensional parameter, but which is otherwise unrestricted. We show that estimation is generally not feasible in this model because of the curse of dimensionality associated with the required estimation of a number of conditional densities. Thus, we develop a multiply-robust framework, in a spirit similar to the approach of Vansteelandt et al (2008), in the sense that our general model assumes at least one of several working models holds. We construct a CAN estimator of β^* under a union semiparametric logistic model that assumes $q_3\left(\mathbf{A},\mathbf{X};\beta\right)$ is correctly specified, and at least one of the following three statements is true:

(i) the models $q_2\left(\mathbf{X},A_2;\gamma_2\right), q_1\left(\mathbf{X},A_1;\gamma_1\right), h(\mathbf{X};\gamma_0)$ are all correct, and thus the working model for the outcome regression of Y on $(\mathbf{A},\mathbf{X})$ is correct,

(ii) the model $q_1\left(\mathbf{X},A_1;\gamma_1\right)$ and a model $f\left(A_1|Y=0,A_2,\mathbf{X};\alpha_1\right)$ for the density $f\left(A_1|Y=0,A_2,\mathbf{X}\right)$ of A_1 given A_2 and $\mathbf{X}$ in unaffected individuals with $Y=0$ are both correct, or

(iii) the model $q_2(\mathbf{X}, A_2; \gamma_2)$ and a model $f(A_2|Y=0, A_1, \mathbf{X}; \alpha_2)$ for the density $f(A_2|Y=0, A_1, \mathbf{X})$ of A_2 given A_1 and $\mathbf{X}$ in unaffected individuals with $Y=0$ are both correct.

Thus, the proposed approach is triply robust as only one of (i)-(iii) needs to hold to obtain a CAN estimator of β^*. A subtle and notable difference with the multiple-robust approach of Vansteelandt et al (2008) lies in the fact that, whereas multiple robust estimation of an additive/multiplicative interaction involves models for $f(A_1|A_2, \mathbf{X})$ and $f(A_2|A_1, \mathbf{X})$, estimation of an odds ratio interaction instead posits models for the retrospective densities $f(A_1|Y=0, A_2, \mathbf{X})$ and $f(A_2|Y=0, A_1, \mathbf{X})$. As we later formalize in this paper, this subtle distinction is key to obtaining multiple robust estimators of an odds ratio interaction.

The paper is organized as follows. The semiparametric logistic model of interaction is introduced in Section 2. The model parameterizes the statistical interaction between exposures A_1 and A_2 (on the logistic scale) as a function of the exposures, and of $\mathbf{X}$, in terms of a finite number of parameters, but leaves the observed data law otherwise unrestricted. In particular, the proposed model leaves 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 asymptotically distribution free test for statistical interaction exists with guaranteed performance in finite samples because estimation of the interaction parameters requires the auxiliary estimation of conditional densities 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 estimator described above. We illustrate the performance of our estimator in a simulation study in Section 5 and the analysis of a case-control cancer study in Section 6.

2 Semiparametric model and inference

Suppose data $(Y_i, \mathbf{A}_i, \mathbf{X}_i)$ is collected for each of $i = 1, ..., n$ independent subjects. Here, Y_i is the binary 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 the logistic scale.

To determine whether an asymptotic distribution-free (ADF) test of the null hypothesis $\beta^* = \mathbf{0}$ is available, we consider the semiparametric interaction model $\mathcal{A}$ in which some of the parametric restrictions of model (2) are relaxed. Specifically, model $\mathcal{A}$ is defined by the conditional mean model

$$\operatorname{logit} \Pr\left(Y = 1|\mathbf{A}, \mathbf{X}\right) = m(\mathbf{A}, \mathbf{X}; \beta^*) \tag{3}$$

where

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

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

$$\operatorname{logit} \Pr\left(Y = 1|\mathbf{A}, \mathbf{X}\right) = \beta^* A_1 A_2 + q_2\left(\mathbf{X}, A_2\right) + q_1\left(\mathbf{X}, A_1\right) + h(\mathbf{X})$$

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

Theorem 1 gives the influence functions of regular asymptotically linear (RAL) estimators of β^* in model $\mathcal{A}$ and will form the basis of our argument as to why estimation of β^* in model $\mathcal{A}$ is infeasible when $\mathbf{X}$ is high dimensional.

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

$$E\{\sigma^2_{Y|A,X}(\mathbf{A}, \mathbf{X})\mathbf{d}(\mathbf{A}, \mathbf{X})|A_1, \mathbf{X}\} = E\{\sigma^2_{Y|A,X}(\mathbf{A}, \mathbf{X})\mathbf{d}(\mathbf{A}, \mathbf{X})|A_2, \mathbf{X}\} = \mathbf{0}, \tag{4}$$

where

$$\sigma^2_{Y|A,X}(\mathbf{A}, \mathbf{X}) = \Pr(Y = 1|\mathbf{A}, \mathbf{X})(1 - \Pr(Y = 1|\mathbf{A}, \mathbf{X}))$$

such that $\widehat{\beta}$ has influence function $d(A, X)\epsilon(\beta)$, where $\epsilon(\beta) = Y - B(A, X; \beta)$ and

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

That is, $n^{1/2}\left(\widehat{\beta} - \beta^*\right) = n^{-1/2} \sum_{i=1}^n d(A_i, X_i)\epsilon_i(\beta) + o_p(1)$.

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

$$\sum_{i=1}^n \mathbf{d}(\mathbf{A}_i, \mathbf{X}_i)\epsilon_i(\beta) = \mathbf{0}, \tag{5}$$

for some $\mathbf{d} \in \mathcal{D}$. The solution $\widetilde{\beta}(\mathbf{d})$ to this equation is an infeasible estimator as the set of functions $\mathbf{D}$ satisfying (4) depends on the unknown conditional law $f(\mathbf{A}_i|\mathbf{X}_i)$ of exposure $\mathbf{A}_i$, given $\mathbf{X}_i$, and $\epsilon_i(\beta)$ 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 a subset of these unknown functions can be consistently estimated. While smoothing methods could in principle be used, with the sample sizes found in practice, the data available to estimate either the density $f(\mathbf{A}_i|\mathbf{X}_i)$ or $q_2(\mathbf{X}_i, A_{i2}), q_1(\mathbf{X}_i, A_{i1})$ and $h(\mathbf{X}_i)$ will be sparse when $\mathbf{X}_i$ is a vector with more than two continuous components. As a consequence any feasible estimator of β^* under model $\mathcal{A}$ will exhibit poor finite sample performance when the predictor space is large. 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 to estimate the unknown auxiliary functions $q_2(\mathbf{X}_i, A_{i2}), q_1(\mathbf{X}_i, A_{i1})$, $h(\mathbf{X}_i)$ and $f(\mathbf{A}_i|\mathbf{X}_i)$.

Before we proceed, we give a key result due to Chen (2007), who established that in the semi-parametric model $\mathcal{A}$ characterized by the sole restriction (3), the density $f(Y_i, \mathbf{A}_i|\mathbf{X}_i)$ can be written as $f(Y_i, \mathbf{A}_i|\mathbf{X}_i; \beta^*)$ where $f(Y, \mathbf{A}|\mathbf{X}; \beta) =$

$$\frac{\mathbf{OR}_{Y,A|X}(Y, \mathbf{A}|\mathbf{X}; \beta) f(Y|\mathbf{A} = \mathbf{0}, \mathbf{X}) f(\mathbf{A}|Y = 0, X)}{\int \mathbf{OR}_{Y,A|X}(Y, \mathbf{A}|\mathbf{X}; \beta) f(y|\mathbf{A} = \mathbf{0}, \mathbf{X}) f(\mathbf{a}|Y = 0, X)\, d\mu(a, y)}, \tag{6}$$

$$f(\mathbf{A}|\mathbf{X}; \beta) = f(1, \mathbf{A}|\mathbf{X}; \beta) + f(0, \mathbf{A}|\mathbf{X}; \beta),$$

with $\mathbf{OR}_{Y,\mathbf{A}|\mathbf{X}}(Y, \mathbf{A}|\mathbf{X}; \beta)$ the conditional odds ratio function relating Y and $\mathbf{A}$ within levels of $\mathbf{X}$:

$$\log \mathbf{OR}_{Y,\mathbf{A}|\mathbf{X}}(Y, \mathbf{A}|\mathbf{X}; \beta) = \{q_3(\mathbf{A}, \mathbf{X}; \beta) + q_2(\mathbf{X}, A_2) + q_1(\mathbf{X}, A_1)\} Y,$$

$$\text{logit} f(Y = 1|\mathbf{A} = \mathbf{0}, \mathbf{X}) = h(\mathbf{X})$$

and $f(\mathbf{A}|Y = 0, X) =$

$$\frac{\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1, A_2|Y = 0, \mathbf{X}) f(A_1|Y = 0, A_2 = 0, \mathbf{X}) f(A_2|Y = 0, A_1 = 0, \mathbf{X})}{\int \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(a_1, a_2|Y = 0, \mathbf{X}) f(A_1 = a_1|Y = 0, A_2 = 0, \mathbf{X}) f(A_2 = a_2|Y = 0, A_1 = 0, \mathbf{X})\, d\mu(a_1, a_2)} \tag{7}$$

$\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1, A_2|Y = 0, \mathbf{X})$ is the unknown conditional odds ratio function relating A_1 and A_2 given $Y = 0$ and $\mathbf{X}$, and $f(A_1|Y = 0, A_2 = 0, \mathbf{X})$ and $f(A_2|Y = 0, A_1 = 0, \mathbf{X})$ are unknown conditional densities which are solely restricted by

$\int \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}\left(a_1,a_2|Y=0,\mathbf{X}\right) f\left(A_1=a_1|Y=0,A_2=0,\mathbf{X}\right) f\left(A_2=a_2|Y=0,A_1=0,\mathbf{X}\right) d\mu\left(a\right) < \infty$, $a.e$

and

$$\int \mathbf{OR}_{Y,\mathbf{A}|\mathbf{X}}\left(y,\mathbf{a}|\mathbf{X};\beta\right) f(y|\mathbf{A}=\mathbf{0},\mathbf{X}) f\left(\mathbf{a}|Y=0,X\right) d\mu\left(a,y\right) < \infty,\ a.e$$

In the following two sections, we demonstrate that multiply robust estimators of β^* are obtained when the parameters of these models are estimated in an appropriate fashion. In Section 3, we assume that A_1 has finite support. This assumption is dropped in Section 4.

3 Polytomous A_1

Suppose A_1 has support $\{0, z_1, ...z_J\}$ and define the vector $(I(A_1 = z_1), ..., I(A_1 = z_J))$ which for convenience, we again denote A_1. There are several important settings in which A_1 is a polytomous factor. For instance, in the context of a genetic study, $J = 2$ with $A_1 \in \{0, z_{1,} z_2\} = \{0, 1, 2\}$ typically corresponds to the three ordered levels of a Single Nucleotide Polymorphism (SNP).

For A_1 polytomous, let $\Psi\left(A_2,\mathbf{X};\beta^*\right) = E\left\{\sigma^2_{Y|A,X}(\mathbf{A},\mathbf{X})\Delta\left(\beta^*\right)^{\otimes 2}|A_2,\mathbf{X}\right\}$ and

$$\mathbf{k} \mapsto V(\beta;\mathbf{k}) = [\mathbf{k}(A_2,\mathbf{X}) - \widetilde{E}\{\mathbf{k}(A_2,\mathbf{X})|\mathbf{X};\beta\}] \times \Delta(\beta)$$

be a function that maps the space of $p-$dimensional functions of A_2 and $\mathbf{X}$ into L_2, where $\widetilde{E}\left\{\mathbf{k}\left(A_2,\mathbf{X}\right)|\mathbf{X};\beta\right\}$

$$= E\{\mathbf{k}(A_2,\mathbf{X}) \times\ \Psi\left(\mathbf{A}_2,\mathbf{X};\beta\right)|\mathbf{X}\} \times E\{\ \Psi\left(\mathbf{A}_2,\mathbf{X};\beta\right)|\mathbf{X}\}^{-1}$$

and

$$\Delta(\beta) = A_1 - E(\sigma^2_{Y|A,X}(\mathbf{A},\mathbf{X})A_1|A_2,\mathbf{X};\beta^*) \times E\left\{\sigma^2_{Y|A,X}(\mathbf{A},\mathbf{X})|A_2,\mathbf{X}\right\}^{-1}$$

L_2 is the Hilbert space of functions of $(Y,\mathbf{A},\mathbf{X})$ with finite variance. The following lemma is proved in the appendix.

Lemma 1: Under model $\mathcal{A}$ with polytomous A_1, the set of estimating equations (5) with $d \in \mathcal{D}$ can equivalently be rewritten as

$$0 = \sum_{i=1}^{n} V_i(\beta;\mathbf{k}) \times \epsilon_i(\beta)$$

where $\mathbf{k} = \mathbf{k}(A_2,\mathbf{X})$ is a member of the set of $p \times K$ functions of $(A_2,\mathbf{X})$. For reasons previously discussed, the solution $\widetilde{\beta}(\mathbf{d})$ to this equation remains an infeasible estimator.

We consider three modeling strategies. 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 $\gamma \equiv (\gamma_0', \gamma_1^{*\prime}, \gamma_2^{*\prime})'$ unknown finite dimensional parameters, and with γ_0, γ_1 and γ_2 variation independent. For the second and third strategies, we postulate models $\mathcal{M}_{y,a_1}$ and $\mathcal{M}_{y,a_2}$ respectively which we describe next. Both of these models share a model for the odds ratio function $\mathbf{or}(A_1,A_2|Y=0,\mathbf{X})$ relating A_1 and A_2 given $\mathbf{X}$ among unaffected individuals $Y=0$,

$$\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1,A_2|Y=0,\mathbf{X}) = \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1,A_2|Y=0,\mathbf{X};\alpha_0^*)$$

where $\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1,A_2|Y=0,\mathbf{X};\alpha_0)$ is a known function smooth in α_0 that satisfies

$$1 = \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(0,A_2|Y=0,\mathbf{X};\alpha_0) = \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1,0|Y=0,\mathbf{X};\alpha_0)$$
$$= \mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}(A_1,A_2|Y=0,\mathbf{X};\mathbf{0})$$

In the second strategy, we also assume $q_1(\mathbf{X},A_1;\gamma_1^*)$ is correct and we further assume a parametric model for the density of A_1 given $\mathbf{X}$ among unaffected individuals $Y=0$ with $A_2=0$,

$$f\left(A_1|Y=0,A_2=0,\mathbf{X}\right)=f\left(A_1|Y=0,A_2=0,\mathbf{X};\alpha_1^*\right)$$

These three assumptions constitute $\mathcal{M}_{y,a_1}$. For the third strategy, we assume $q_2\left(\mathbf{X},A_2\right)=q_2\left(\mathbf{X},A_2;\gamma_2^*\right)$ is correct and further assume a parametric model for the density of A_2 given $\mathbf{X}$ among unaffected individuals $Y=0$ with $A_1=0$,

$$f\left(A_2|Y=0,A_1=0,\mathbf{X}\right)=f\left(A_2|Y=0,A_1=0,\mathbf{X};\alpha_2^*\right)$$

Together with $\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}\left(A_1,A_2|Y=0,\mathbf{X};\alpha_0\right)$ these last two models constitute $\mathcal{M}_{y,a_2}$. Since we cannot be certain that any of the three models $\mathcal{M}_y,\mathcal{M}_{y,a_1}$ or $\mathcal{M}_{y,a_2}$ is 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 $\mathcal{B}\equiv\mathcal{A}\cap\left(\mathcal{M}_y\cup\mathcal{M}_{ya_1}\cup\mathcal{M}_{ya_2}\right)$ of model $\mathcal{A}$ that assumes that at least one of $\mathcal{M}_y,\mathcal{M}_{ya_1}$, and $\mathcal{M}_{ya_2}$ is true. In line with Robins and Rotnitzky (2001) and Vansteelandt et al, (2008), we refer to such estimators as multiply robust estimators. Part (i) of Theorem 2 below shows that, under mild regularity conditions, the estimator $\widehat{\beta}\equiv\widehat{\beta}\left(\mathbf{k}\right)$ is multiply robust, in the sense of being CAN for β^* under model $\mathcal{B}$, for $\widehat{\beta}\left(\mathbf{k}\right)$ the solution to

$$\begin{aligned}0&=\sum_{i=1}^{n}\mathbf{U}_i\left(\widehat{\beta},\widehat{\gamma}\left(\widehat{\beta}\right),\widehat{\alpha}\left(\widehat{\beta}\right);\mathbf{k}\right)\\&=\sum_{i=1}^{n}\left[\mathbf{k}(A_{2i},\mathbf{X}_i)-\widetilde{E}\{\mathbf{k}(A_{2i},\mathbf{X}_i)|\mathbf{X}_i;\widehat{\beta},\widehat{\gamma}\left(\widehat{\beta}\right),\widehat{\alpha}\left(\widehat{\beta}\right)\}\right]\\&\quad\times\Delta_i(\left(\widehat{\beta},\widehat{\gamma}\left(\widehat{\beta}\right),\widehat{\alpha}\left(\widehat{\beta}\right)\right)\times\epsilon_i\left(\widehat{\beta},\widehat{\gamma}\left(\widehat{\beta}\right)\right)\end{aligned}\tag{8}$$

with $\epsilon_i(\widehat{\beta},\widehat{\gamma}\left(\widehat{\beta}\right)) = Y - B\left(A, X; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right)$, $\mathbf{k}(A_{2i}, \mathbf{X}_i)$ an arbitrary $p \times J$ function of $(A_{2i}, \mathbf{X}_i)$, for $\widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right)$, $\widehat{\alpha}_2\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right), \widehat{\gamma}_2\left(\widehat{\beta}\right)$ and $\widehat{\gamma}_0\left(\widehat{\beta}\right)$ solving the system of equations:

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{G}_{0i}(\widehat{\alpha}\left(\widehat{\beta}\right), \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\beta}) \tag{9}$$

$$\equiv \sum_{i=1}^{n} \left[\mathbf{c}_0(A_{2i}, \mathbf{X}_i) - \widetilde{E}\{\mathbf{c}_0(A_{2i}, \mathbf{X}_i)|Y_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\alpha}\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right), \widehat{\gamma}_2\left(\widehat{\beta}\right)\}\right]$$

$$\times \left[A_{1i} - E\left\{A_{1i}|A_{2i}, Y_i, \mathbf{X}_i, Y_i, \widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right)\right\}\right]$$

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{G}_{1i}\left(\widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right)\right) \tag{10}$$

$$\equiv \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_1} \ln f\left(A_{1i}|A_{2i}Y_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \alpha_1, \widehat{\gamma}_1\left(\widehat{\beta}\right)\right)_{|\alpha_1 = \widehat{\alpha}_1\left(\widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right)\right)}$$

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{G}_{2i}\left(\widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right)\right) \tag{11}$$

$$\equiv \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_2} \ln f\left(A_{2i}|A_{1i}Y_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \alpha_2, \widehat{\gamma}_2\left(\widehat{\beta}\right)\right)_{|\alpha_2 = \widehat{\alpha}_2\left(\widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\gamma}_2\left(\widehat{\beta}\right)\right)}$$

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{H}_{1i}\left(\beta, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right)\right) \tag{12}$$

$$\equiv \sum_{i=1}^{n} \left[\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i) - \frac{E\left[\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i) Var\left\{Y_i|\mathbf{A}_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right\}|A_{2i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}\left(\widehat{\beta}\right)\right]}{E\left[Var\left\{(Y_i|\mathbf{A}_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right\}|A_{2i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}\left(\widehat{\beta}\right)\right]}\right] \epsilon_i\left(\widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right)$$

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{H}_{2i}\left(\beta, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right)\right) \tag{13}$$

$$\equiv \sum_{i=1}^{n} \left[\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i) - \frac{E\left[\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i) Var\left\{Y_i|\mathbf{A}_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right\}|A_{1i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}\left(\widehat{\beta}\right)\right]}{E\left[Var\left\{(Y_i|\mathbf{A}_i, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right\}|A_{1i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}\left(\widehat{\beta}\right)\right]}\right] \epsilon_i\left(\widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right)$$

$$\mathbf{0} = \sum_{i=1}^{n} \mathbf{H}_{3i}\left(\beta, \widehat{\gamma}\left(\widehat{\beta}\right), \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right)\right) \tag{14}$$

$$= \sum_{i=1}^{n} \mathbf{c}_3(\mathbf{X}_i)\epsilon_i(\widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right))$$

for arbitrary vector functions $\mathbf{c}_0(A_{2i}, \mathbf{X}_i)$, $\mathbf{c}_1(\mathbf{A}_i, \mathbf{X}_i)$, $\mathbf{c}_2(\mathbf{A}_i, \mathbf{X}_i)$ and $\mathbf{c}_3(\mathbf{X}_i)$ of the dimension of α_0, γ_1, γ_2 and γ_0, respectively. The arguments of Robins and Rotnitzky (2001) imply that a necessary condition for the existence of such a triply robust estimator of β^* in model $\mathcal{A} \cap (\mathcal{M}_y \cup \mathcal{M}_{ya_1} \cup \mathcal{M}_{ya_2})$ is that there exist an unbiased estimating equation for β^* (with non-trivial power against local alternatives) when any of the following three statements holds: (1) $q_2(\mathbf{X}, A_2), q_1(\mathbf{X}, A_1)$ and $h(\mathbf{X})$ are all known, (2) $q_1(\mathbf{X}, A_1)$, and $f(A_1|A_2, Y, \mathbf{X})$ are both known (3) $q_2(\mathbf{X}, A_2)$, and $f(A_2|A_1, Y, \mathbf{X})$ are both known, The main step in the proof of Theorem 2 is showing that, for $j = 1, 2, 3$, (8) is an unbiased estimating equation for β^* 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 of the theorem is then completed by showing that all of the following are true:$B\left(A, X; \widehat{\beta}, \widehat{\gamma}\left(\widehat{\beta}\right)\right)$ is a CAN estimator of $f(Y = 1|\mathbf{A}, \mathbf{X})$ in model $\mathcal{M}_y$, $f\left(A_{1i}|Y_i, A_{2i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_1\left(\widehat{\beta}\right), \widehat{\gamma}_1\left(\widehat{\beta}\right)\right)$ is a CAN estimator of $f(A_{1i}|Y_i, A_{2i}, \mathbf{X}_i)$ in model $\mathcal{M}_{ya_1}$,and $f\left(A_{2i}|Y_i, A_{1i}, \mathbf{X}_i; \widehat{\beta}, \widehat{\alpha}_0\left(\widehat{\beta}\right), \widehat{\alpha}_2\left(\widehat{\beta}\right), \widehat{\gamma}_2\left(\widehat{\beta}\right)\right)$ is a CAN estimator of $f(A_{1i}|Y_i, A_{2i}, \mathbf{X}_i)$ in model $\mathcal{M}_{ya_2}$.

Theorem 2. Suppose that the regularity conditions stated in the appendix hold and that $\beta, \alpha_0, \alpha_1, \alpha_2, \gamma_0', \gamma_1^{*\prime}$, and γ_2^* are variation independent.

(i) Then, $\sqrt{n}(\widehat{\beta} - \beta^*)$ is RAL under model $\mathcal{B}$ with influence function

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

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

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

with $\tilde{\theta}(\beta^*) = (\tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))$ denoting the probability limits of the estimators $\widehat{\theta}\left(\widehat{\beta}\right) = (\widehat{\gamma}(\widehat{\beta}), \widehat{\alpha}(\widehat{\beta}))$, *respectively, and*

$$\mathbf{U}_i^*(\beta,\theta;\mathbf{k}) = \mathbf{U}_i(\beta,\theta;\mathbf{k}) - E\left\{\frac{\partial}{\partial\theta}\mathbf{U}_i(\beta,\theta;\mathbf{k})\right\}E^{-1}\left\{\frac{\partial}{\partial\theta}\mathbf{R}_i(\beta,\theta)\right\} \times \mathbf{R}_i(\beta,\theta)$$

with $\mathbf{R}_i(\beta,\theta) \equiv (\mathbf{H}_i'(\beta,\theta), \mathbf{G}_i'(\beta,\theta))'$, $\mathbf{H}_i(\beta,\theta) \equiv (\mathbf{H}_{i1}'(\beta,\theta), \mathbf{H}_{i2}'(\beta,\theta), \mathbf{H}_{i3}'(\beta,\theta))'$ *and* $\mathbf{G}_i(\beta,\theta) \equiv (\mathbf{G}_{i0}'(\beta,\theta),\ \mathbf{G}_{i1}'(\beta,\theta),\ \mathbf{G}_{i2}'(\beta,\theta))'$.

(ii) Furthermore, let $\widehat{\beta}(\mathbf{k},\mathbf{G}_{(1)},\mathbf{H}_{(1)})$ *and* $\widehat{\beta}(\mathbf{k},\mathbf{G}_{(2)},\mathbf{H}_{(2)})$ *be 2 estimators of* β^* *under model* $\mathcal{B}$ *corresponding to the same index functions* $\mathbf{k}$*, but different unbiased estimating functions* $\mathbf{G}_{(1)}$ *and* $\mathbf{G}_{(2)}$ *for* α_0^* *under model* $\mathcal{A}\cap(\mathcal{M}_{ya_1}\cup\mathcal{M}_{ya_2})$, α_1^* *under model* $\mathcal{A}\cap\mathcal{M}_{ya_1}$, α_2^* *under model* $\mathcal{A}\cap\mathcal{M}_{ya_2}$;*and* $\mathbf{H}_{(1)}$ *and* $\mathbf{H}_{(2)}$ *for* γ_0^* *under model* $\mathcal{A}\cap\mathcal{M}_y$, γ_1^* *under model* $\mathcal{A}\cap(\mathcal{M}_y\cup\mathcal{M}_{ya_1})$, and γ_2^* *under model* $\mathcal{A}\cap(\mathcal{M}_y\cup\mathcal{M}_{ya_2})$. *Then,* $\sqrt{n}\left\{\widehat{\beta}(\mathbf{G}_{(1)},\mathbf{H}_{(1)}) - \widehat{\beta}(\mathbf{G}_{(2)},\mathbf{H}_{(2)})\right\} = o_p(1)$ *at the intersection model* $\mathcal{A}\cap\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap\mathcal{M}_{ya_2}$.

(iii) For any choice of $(\mathbf{G},\mathbf{H})$, $\widehat{\beta}(\widehat{\mathbf{k}}_{opt})$ *is semiparametric locally efficient in the sense that it is RAL under model* $\mathcal{B}$ *and achieves the semiparametric efficiency bound for* $\mathcal{B}$ *at the intersection submodel* $\mathcal{A}\cap\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap\mathcal{M}_{ya_2}$; where

$$\widehat{\mathbf{k}}_{opt} = \Omega\left(A_2,\mathbf{X};\beta^\dagger\right),$$

$\Omega\left(A_2,\mathbf{X};\beta\right)$ *is the* $p\times K$ *matrix with kth column given by* $\partial q_3\left(A_1 = z_k, A_2, \mathbf{X};\beta\right)/\partial\beta$, *and* $\beta^\dagger$ *is a preliminary estimator which is CAN at the intersection submodel*

Part (i) of Theorem 2 suggests that multiply robust estimators of β^* in model $\mathcal{B}$ 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 $\mathcal{B}$ has the same asymptotic distribution as $\widehat{\beta}\left(\mathbf{k}\right)$ for some $\mathbf{k}$ and, thus, that any multiply robust estimator in model

$\mathcal{B}$ can be obtained in this way. An empirical version of Σ can be used to estimate the large sample covariance of $\widehat{\beta}(\mathbf{k})$, or alternatively, a bootstrap estimator can also be used. Part (ii) of Theorem 2 suggests that the choice of estimators for γ^* and α_0^* has no impact on the efficiency of $\widehat{\beta}$ when the models $\mathcal{M}_y, \mathcal{M}_{ya_1}$ and $\mathcal{M}_{ya_2}$ are correctly specified. Thus the fact that γ_1^*, γ_2^* and α_0^* are estimated by solving equations (12), (13) and (9), respectively, rather than by the more efficient maximum likelihood estimators under say model model $\mathcal{A}\cap\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap\mathcal{M}_{ya_2}$ has no effect on the asymptotic variance of $\widehat{\beta}$ when the law of the data lies in the intersection submodel$\mathcal{A}\cap\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap\mathcal{M}_{ya_2}$. Nonetheless, the use of these specific estimating equations is critical to control bias. Indeed, (12), (13) and (9) are the doubly robust estimating equations developed by Tchetgen Tchetgen, Robins and Rotnitzky (2010) that are guaranteed to yield CAN estimators of γ_1^* , γ_2^* and α_0^* under the semiparametric odds ratio models $\mathcal{A}\cap(\mathcal{M}_y\cup\mathcal{M}_{ya_1}), \mathcal{A}\cap(\mathcal{M}_y\cup\mathcal{M}_{ya_2})$ and $\mathcal{A}\cap(\mathcal{M}_{ya_1}\cup\mathcal{M}_{ya_2})$ respectively. To be more specific, consider equation (12) which is double robust for (γ_1^*, β^*) under model $\mathcal{A}\cap(\mathcal{M}_y\cup\mathcal{M}_{ya_1})$ that assumes the odds ratio model

$$\mathbf{OR}_{Y,A_1|A_2,\mathbf{X}}\left(Y, A_1|A_2, \mathbf{X};\ \beta^*, \gamma_1^*\right) = \exp\left\{q_3\left(\mathbf{A}, \mathbf{X}; \beta^*\right)Y + q_1\left(\mathbf{X}, A_1; \gamma_1^*\right)Y\right\}$$

is correct and either model $\mathcal{M}_y$ is correct and thus $f\left(Y|A_1=0, A_2, L\right)$ is correctly modeled, or model $\mathcal{M}_{ya_1}$ holds and thus $f\left(A_1|Y=0, A_2, L\right)$ is correctly modeled but not necessarily both. As demonstrated in the proof of Theorem 2, it is precisely our careful use of these doubly robust estimators (instead of maximum likelihood estimators) that makes our multiply robust approach possible.

Part (iii) of the theorem gives a locally efficient estimator of β^* under model $\mathcal{B}$ at the intersection submodel $\mathcal{A}\cap\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap\mathcal{M}_{ya_2}$. A theorem due to Robins and Rotnitzky (2001) implies the semiparametric variance bound in models $\mathcal{B}$ and $\mathcal{A}$ coincide whenever the model $\mathcal{M}_y\cap\mathcal{M}_{ya_1}\cap$

$\mathcal{M}_{ya_2}$ is true, and thus $\widehat{\beta}(\widehat{\mathbf{k}}_{opt})$ is semiparametric efficient in model $\mathcal{A}$ at the intersection model $\mathcal{M}_y \cap \mathcal{M}_{ya_1} \cap \mathcal{M}_{ya_2}$.

4 Continuous A

In this section, we generalize the results from the previous section and consider the setting where $\mathbf{A}$ is continuous. To proceed, we need the following definition from Tchetgen Tchetgen, Robins and Rotnitzky (2010):

Definition: Admissible Independence Density: Given conditional densities $f^{\dagger}(Y|\mathbf{X}), g_1^{\dagger}(A_1|\mathbf{X})$ *and* $g_2^{\dagger}(A_2|\mathbf{X})$*, the density* $h^{\dagger}(\mathbf{A}|\mathbf{X}) = g_1^{\dagger}(A_1|\mathbf{X})\, g_2^{\dagger}(A_2|\mathbf{X})$ *that makes* A_1 and A_2 *conditionally independent given* $\mathbf{X}$ *is an admissible independence density if the joint law of* $\mathbf{A}$ *given* $\mathbf{X}$ *under* $h^{\dagger}(\mathbf{A}|\mathbf{X})$ *is absolutely continuous wrt to the true law of* $\mathbf{A}$ *given* $\mathbf{X}$ *with probability one. Furthermore,* $E^{\dagger}(\cdot|\cdot, \mathbf{X})$ *denotes conditional expectations with respect to* $h^{\dagger}(\mathbf{A}|\mathbf{X})$.

The following lemma is proved in the appendix:

Lemma 2: Given an admissible independence density $h^{\dagger}$*, under model* $\mathcal{A}$*, the set of estimating equations (5) with* $d \in D$ *can equivalently be rewritten as*

$$0 = \sum_{i=1}^{n} \{\mathbf{d}(\mathbf{A}_i, \mathbf{X}_i) - \mathbf{d}^{\dagger}(\mathbf{A}_i, \mathbf{X}_i)\} \frac{h^{\dagger}(\mathbf{A}_i|\mathbf{X}_i)}{f(Y_i, \mathbf{A}_i|\mathbf{X}_i; \beta^*)} (-1)^{1-Y_i}$$

with $f(Y_i, \mathbf{A}_i|\mathbf{X}_i; \beta^*)$ *defined in Eq.* (6) *and* $d^{\dagger}(\mathbf{A}, \mathbf{X}) = E^{\dagger}(\mathbf{D}|A_1, \mathbf{X}) + E^{\dagger}(\mathbf{D}|A_2, \mathbf{X}) - E^{\dagger}(\mathbf{D}|\mathbf{X})$*.with* $\mathbf{d} \in \mathbf{D}$ *a member of the set of* $p \times 1$ *functions of* (A, X_i).

A multiply-robust estimator of the interaction parameter β^* is then obtained as in the previous section, with $\mathbf{U}_i(\beta, \alpha, \gamma; \mathbf{d}, h^{\dagger})$ obtained upon substituting $f(\mathbf{A}_i, Y_i|\mathbf{X}_i; \beta, \gamma, \alpha)$ for $f(Y, \mathbf{A}_i|\mathbf{X}_i; \beta^*)$

in the estimating function given in the above display,and by replacing $\mathbf{G}_{0i}(\alpha,\gamma,\beta)$ with

$$\left\{c_0\left(\mathbf{A}_i,\mathbf{X}_i\right)-c_0^{\dagger}\left(\mathbf{A}_i,\mathbf{X}_i\right)\right\}\frac{h^{\dagger}\left(\mathbf{A}_i|\mathbf{X}_i\right)}{f\left(\mathbf{A}_i|Y_i,\mathbf{X}_i;\beta,\gamma,\alpha\right)}$$

in equation (9) for estimating α_0^*; where $c_0\left(\mathbf{A}_i,\mathbf{X}_i\right)$ is a user-specified function of $\mathbf{A}_i$ and $\mathbf{X}_i$,of the dimension of α_0^*. Parts (i) and (ii) of Theorem 2 can then be shown to still hold, even when, as will be convenient in practice, $g_1^{\dagger}$ and $g_2^{\dagger}$ and thus $h^{\dagger}=g_1^{\dagger}\times g_2^{\dagger}$ is estimated from the observed data. Specifically, the asymptotic distribution of $\widehat{\beta}=\widehat{\beta}\left(\mathbf{d};\widehat{h}^{\dagger}\right)$ thus obtained, is equal to that of $\widehat{\beta}\left(d;h^{\dagger}\right)$ given in Theorem 2 upon making the aforementioned substitutions, with $h^{\dagger}$ the probability limit of $\widehat{f}^{\dagger}\times\widehat{g}^{\dagger}$. Unfortunately, Part (iii) of Theorem 2 only applies when as in the previous section, either A_1 or A_2 has finite support. If in fact A_1 and A_2 are both continuous, the efficient score of β^* is not available in closed form, and thus locally efficient estimation is not possible using this approach. Instead, we undertake an alternative approach similar to Tchetgen Tchetgen et al (2010). The approach is based on a result due to Newey (1993). We take a basis system $\phi_s\left(\mathbf{A},\mathbf{X}\right)$, s= $1,...$ of functions dense in the Hilbert space L_2 of functions of $\mathbf{A},\mathbf{X}$ with finite variance (e.g. tensor products of trigonometric, wavelets or polynomial bases when the components of $\mathbf{A},\mathbf{X}$ are all continuous). For some finite $S>\dim\left(\beta^*\right)$, we form the $S-$dimensional vector $\mathbf{U}\{\beta,\gamma,\alpha;\widetilde{\phi}_K,\widehat{h}^{\dagger}\}$ with $\widetilde{\phi}_S$ the vector of the first S basis functions and let $\widehat{W}_S\left(\beta\right)\equiv\mathbf{U}\{\beta,\widehat{\gamma}\left(\beta\right),\widehat{\alpha}\left(\beta\right);\widetilde{\phi}_S,\widehat{h}^{\dagger}\}$, and $\widehat{\Gamma}_S\left(\beta^{\dagger}\right)=\sum_{i=1}^n\widehat{W}_{S,i}\left(\beta^{\dagger}\right)\widehat{W}_{S,i}^T\left(\beta^{\dagger}\right)$, where $\beta^{\dagger}$ is any preliminary doubly robust estimator of β^*. Let $\widehat{\beta}_{S,\mathrm{eff}}\equiv\widehat{\beta}_{S,\mathrm{eff}}\left(\widetilde{\phi}_S,\widehat{h}^{\dagger}\right)$ be the minimizer of the quadratic form $\left\{\sum_{i=1}^n\widehat{W}_{S,i}\left(\beta\right)\right\}^T\left\{\widehat{\Gamma}_S\left(\beta^{\dagger}\right)\right\}^{-}\left\{\sum_{i=1}^n\widehat{W}_{S,i}\left(\beta\right)\right\}$ with $\left\{\widehat{\Gamma}_S\left(\beta^{\dagger}\right)\right\}^{-}$ a generalized inverse of $\widehat{\Gamma}_K\left(\beta^{\dagger}\right)$. Then, $\widehat{\beta}_{S,\mathrm{eff}}\equiv\widehat{\beta}_{S,\mathrm{eff}}\left(\widetilde{\phi}_S,\widehat{h}^{\dagger}\right)$ is consistent and asymptotically normal in the semi-parametric union model $\mathcal{B}$; furthermore with S chosen sufficiently large, the asymptotic variance of $n^{1/2}(\widehat{\beta}_{S,\mathrm{eff}}-\beta^*)$ nearly attains the semi-parametric efficiency bound for the union model at the intersection sub-model with all working models correct. In particular, the inverse of the asymptotic variance of $\widehat{\beta}_{S,\text{eff}}$ at the intersection sub-model is

$$\begin{aligned}\Theta_S &= \left[E\left\{\frac{\partial}{\partial\beta^T}\mathbf{U}\left(\beta,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)|_{\beta=\beta^*}\right\}\right]^T\Gamma_S^- E\left\{\frac{\partial}{\partial\beta^T}\mathbf{U}\left(\beta,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)|_{\beta=\beta^*}\right\}\\ &= E\left\{\mathbf{S}_\beta^T\mathbf{U}\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)\right\}\Gamma_S^-\left[E\left\{\mathbf{S}_\beta^T\mathbf{U}\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)\right\}\right]^T\end{aligned}$$

where Γ_S^- is a generalized inverse of $\Gamma_S = E\left\{\mathbf{U}^T\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)\mathbf{U}\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)\right\}$. Thus Θ_S is the variance of the population least squares regression of the score of β, $\mathbf{S}_\beta$ on the linear span of $\mathbf{U}\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)$. By $\widetilde{\phi}_K$ dense in L_2, as $S\to\infty$, the components of $\mathbf{U}\left(\beta^*,\gamma^*,\alpha^*;\widetilde{\phi}_S,h^\dagger\right)$ become dense in the orthogonal complement to the nuisance tangent space

$$\Lambda_{\text{nuis}}^\perp = \left\{\mathbf{U}_i\left(\beta^*,\gamma^*,\alpha^*;\mathbf{d},h^\dagger\right):\mathbf{d}\in\mathcal{D}\right\}$$

of $\mathcal{B}$ by Lemma 2, so that $\Theta_S \underset{S\to\infty}{\to} ||\Pi\left(\mathbf{S}_\beta|\Lambda_{\text{nuis}}^\perp\right)||^2 = Var(S_{\beta,\mathbf{eff}})$, the semi-parametric information bound for estimating β^* under model $\mathcal{B}$, with $S_{\beta,\mathbf{eff}}$ the efficient score of β^*.

5 A Simulation Study

We evaluated the finite sample performance of our locally efficient multiply robust estimator of an odds -ratio statistical interaction. Each experiment was based on 500 replications of random samples of size 800 generated as follows. We generated a vector of auxiliary covariates $\mathbf{X} = (X_1, X_2)$ where $X_1 \overset{i.i.d}{\sim} N(0,1)$ and $X_2 \tilde{}\, Bernoulli(1/2)$. We generated a vector of binary variables (Y, A_1, A_2) with joint conditional density given $\mathbf{X}$ given by eqn (6) where we specified :

$m(\mathbf{A},\mathbf{X};\beta^*) = -0.5 + 0.5X_2 - 0.8X_1^2 + 0.75X_2X_1^3 + 0.5A_1 - 0.6A_2 + 0.6A_1A2$

$\text{logit} f\left(A_1 = 1|Y=0, A_2=0,\mathbf{X};\alpha_1\right) = 0.3 - 0.4X_2 + 0.61X_1^2 - 0.71X_1^2X_2$

$\text{logit} f(A_2 = 1 | Y = 0, A_1 = 0, \mathbf{X}; \alpha_2) = 0.2 + 0.4X_2 - 0.65X_1^2 + 0.4X_1^2 X_2$

$\log \mathbf{OR}_{A_1, A_2 | Y=0, \mathbf{X}} (A_1, A_2 | Y = 0, \mathbf{X}) = -0.4A_1 A_2$

Thus $q_3(A, X; \beta^*) = 0.6A_1 A2$. In each simulation experiment, two estimators were calculated under model $\mathcal{A}$ with $q_3(A, X; \beta) = \beta A_1 A2$. The first is an ordinary logistic regression estimate under working model $\mathcal{M}_y$ with $q_2(A_2, X; \gamma_2^*) = \gamma_2^* A_2, q_1(A_1, X; \gamma_1^*) = \gamma_1^* A_1$ and $h(X; \boldsymbol{\gamma}_0^*) = \gamma_{0,0}^* + \gamma_{0,1}^* X$. The second estimator is the new locally efficient multiply robust estimator which yields consistent estimator of γ_1^* in model $\mathcal{B}$.

The results of the simulation study are summarized in Table 1.

The results indicate that as predicted by theory, the multiply robust estimator produces nearly unbiased estimates for the statistical interaction parameter under model $\mathcal{A} \cap \mathcal{M}_y$, model $\mathcal{A} \cap \mathcal{M}_{ya_1}$, and model $\mathcal{A} \cap \mathcal{M}_{ya_2}$ respectively. In comparison, the standard outcome regression approach is substantially biased under misspecification of $\mathcal{M}_y$. The extra-robustness of the proposed approach comes at a cost in terms of efficiency as apparent in Table 1. The efficiency loss is particularly important when the conditional mean model for the outcome is correctly specified, but overall, the proposed semiparametric approach had reasonable efficiency. As expected, both methods are severely biased when $\mathcal{B}$ is misspecified.

6 Data application

In this section, we illustrate the various methods in an analysis of data from a population-based case-control study based on all ovarian cancer patients identified in Israel between 1 March 1994 and 30 June 1999 (Modan et al 2001). Two controls per case were selected from the central population registry matching on age within two years, area of birth and place and length of residence. Blood samples was collected on both cases and controls and used to test for the presence of mutation in two major breast and ovarian cancer susceptibility genes BRCA1 and BRCA2. Additional data was collected on reproductive and gynecological history such as parity, number of years of oral contraceptive use and gynecological surgery. The main objective of the study was to examine the interplay of the BRCA1/2 genes and known reproductive/gynecological risk factors of ovarian cancer. To test for interactions between reproductive risk factors and BRCA1/2 in their effects on the risk of ovarian cancer, the authors tested for a gene-environment interaction using a standard logistic regression analysis. They also performed an unadjusted case-only analysis of interaction (Piegorsch et al. 1994) under an assumption that genetic variants and environment factor are unconditionally independent in the population. Chatterjee and Carroll (2005) and Tchetgen Tchetgen (2011) re-analyzed these data using a fully parametric logistic regression model for disease given the gene and the gene given disease respectively, further conditioning on environmental and confounding factors $\mathbf{X}$, under a weaker conditional independence assumption of gene and environment given the measured covariates $\mathbf{X}$. The estimator of Chatterjee and Carroll (2005) also required a model for the density $[A_1|A_2, \mathbf{X}]$ and thus it may result in biased estimates of interactions if the working genetic model is incorrect or if their specified working model for $[Y|A_1, A_2, \mathbf{X}]$ is incorrect. Furthermore a violation of the independence assumption can also invalidate an inference based on this approah. The estimator of Tchetgen Tchetgen (2011) which assumes gene-environment conditional (on covariates) independence among controls, may result in biased estimates if either the independence assumption fails or the required working model for $[A_1|Y, A_2, \mathbf{X}]$ is false. In contrast, the semiparametric case-only estimator of Tchetgen Tchetgen and Robins (2010) is endowed with a partial protection against model mis-specification of the required model for the association of $\mathbf{X}$ and A_1 among the unexposed cases. Specifically this estimator of interaction remains CAN if A_1 and A_2 are conditionally independent and either a model for $[A_1|A_2 = 0, \mathbf{X}, D = 1]$ or a model for $[A_2|A_1 = 0, \mathbf{X}, D = 1]$ is correct but not necessarily both. However, the latter approach fails to be consistent if the required independence assumption is false or both working models are false.

In our re-analysis, we illustrate the multiply robust method developed in this paper, and thus as in Modan et al (2001) and Chatterjee and Carroll (2005), we hypothesize that gene-environment interactions operate on a logistic scale in the underlying population. Specifically, we used data on 832 cases and 747 controls who did not have bilateral oophorectomy, were interviewed for risk factor information and successfully tested for BRCA1/2 mutations. Our primary aim was to provide multiply robust inference for an interaction between the dichotomous variable representing a person's BRCA1/2 mutation status and her use of oral contraceptives, where the latter was coded as use for over six years vs use for six years or less.

We assume conditional independence of gene and environmental factors (oral contraceptive and parity which we coded as a count of live births with 10 or more births coded as 10) given age (categorical defined by decades), ethnic background (Ashkenazi or non-Ashkenazi), the presence of personal history of breast cancer, a history of gynecological surgery, and family history of breast or ovarian cancer (no cancer vs one breast cancer in the family vs one ovarian cancer or two or more breast cancer cases in the family). The independence assumption was incorporated in our reanalysis by setting $\log\mathbf{OR}_{A_1,A_2|Y=0,\mathbf{X}}\left(A_1,A_2|Y=0,\mathbf{X}\right)=0$, which is approximately correct under a rare disease assumption. Logistic regression models were used for $\text{logit}f\left(A_1=1|Y=0,A_2=0,\mathbf{X};\alpha_1^*\right)=\alpha_1^{*T}\mathbf{X}$ and $\text{logit}f\left(A_2=1|Y=0,A_1=0,\mathbf{X};\alpha_2^*\right)=\alpha_2^{*T}\mathbf{X}$ with main effects for components of $\mathbf{X}$. We specified a working regression model (2), with $A_1=I(BRCA1/2=1)$ and $A_2=I(OC\text{ use}>6\ \ yrs)$, $q_1\left(\mathbf{X},A_1;\gamma_1^*\right)=\gamma_{01}^*A_1+\gamma_{11}^*A_1\times Parity, \gamma_1^*=\left(\gamma_{01}^*,\gamma_{11}^*\right)$; $q_2\left(\mathbf{X},A_2;\gamma_2^*\right)=\gamma_{20}^*A_2+\gamma_{21}^*A_2\times\ Parity$; $h(\mathbf{X};\gamma_0^*)=\gamma_0^{*T}\mathbf{X}$; $q_3\left(\mathbf{A},\mathbf{X};\beta^*\right)=\beta^*A_1A_2$.

We obtained results for standard logistic regression and our multiply robust approach. Both approaches indicated a strong genetic effect among childless women for whom parity=0 and contraceptive use $\leq$ 6 years. Among these women, our reanalysis confirmed the well established association between BRCA1/2 mutation and ovarian cancer, producing a large increase in risk of ovarian cancer ($\widehat{\gamma}_{01,\text{logistic}} = 3.14$ with corresponding weighted bootstrap s.e.=0.37, $\widehat{\gamma}_{01,\text{dr}} = 3.26$, with corresponding weighted bootstrap s.e.= 0.33). We note however that our estimate of the main effect of BRCA1/2 is in theory more reliable since the estimator $\widehat{\gamma}_{01,\text{dr}}$ that solves equation (12) remains valid under any law in model $\mathcal{A} \cap (\mathcal{M}_y \cup \mathcal{M}_{ya_1})$ (Tchetgen Tchetgen et al, 2010), whereas standard logistic regression assumes the smaller model $\mathcal{M}_y$ is correct.The logistic regression approach produced the estimate of an interaction $\widehat{\beta}_{\text{logistic}} = 0.65$ (weighted bootstrap s.e.= 1. 13) whereas the multiply robust approach gave $\widehat{\beta}_{mr} = 0.82$ (weighted bootstrap s.e.= 1. 044). Thus, the multiple robust approach which in theory is less prone to bias, suggests a larger interaction (although not statistically significant) between BRCA1/2 and OC use. The increased precision of the new approach may be due to the $G - E$ independence assumption that we explicitly incorporated into the multiply robust analysis, whereas standard logistic regression is known not to use this assumption.

7 Conclusion

In this article, we have developed a class of multiply robust estimators of an odds ratio statistical interaction. Whereas multiply robust estimators previously proposed to evaluate additive and multiplicative interactions use a model for the density of exposures given covariates (Vansteelandt et al, 2008), in the case of an odds ratio interaction we show that multiply robust inference instead requires a model for the conditional density of the exposures given covariates *amongst noncases* (see Tchetgen Tchetgen et al (2010) for further discussion). An implication of our result is that in contrast with an additive interaction (Vansteelandt et al, 2008), asymptotically distribution-free tests of the no-conditional (on$\mathbf{X}$) -interaction hypothesis are generally not available when high dimensional $\mathbf{X}$ is observed, even in settings where one of the exposure conditional (on $\mathbf{X}$ distribution) is known, as is often the case in randomized follow-up studies and family-based genetic association studies. Nevertheless, we recommend that our proposed approach be used quite generally, because an inference concerning an interaction effect under our approach has multiple chances, rather than only one chance, to be correct or nearly correct. We also note that the approach is easy to extend to more general outcomes, by modifying the methods developed here along the lines of those in Tchetgen Tchetgen et al (2010) to accommodate a non-binary outcome . We conclude by emphasizing that, the well-known invariance property of the odds ratio functional to alterations in the marginal distributions of $(Y, \mathbf{X})$ or of $(A, \mathbf{X})$ implies that our methodology readily applies to both prospective and retrospective study designs and may also be adapted to more general outcome dependent sampling settings.

References

[1] Bickel, P., Klassen, C., Ritov, Y. and Wellner, J. (1993). Efficient and Adaptive Estimation for Semi-parametric Models. Springer, New York.

[2] Chatterjee, N. and R. J. Carroll (2005): "Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies," Biometrika, 92, 399–418.

[3] Yun Chen, H (2007). A Semi-parametric Odds Ratio Model for Measuring Association. *Biometrics*, Volume 63, Number 2, pp. 413-421(9)

[4] Modan, M. D., Hartge, P. et al. (2001). Parity, oral contraceptives and the risk of ovarian cancer among carriers and noncarriers of a BRCA1 or BRCA2 mutation. *New England Journal of Medicine.* 345, 235–40.

[5] Newey, W. (1993), “Efficient Estimation of Models with Conditional Moment Restrictions”, in Maddala, Rao and Vinod (eds) Handbook of Statistics, IV, Ch. 11, Elsevier Science.

[6] Piegorsch, W. W., Weinberg, C. R. & Taylor, J. A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population based case-control studies. Statist. Med. 13, 153–62.

[7] Robins, J. M., Mark, S. D. and Newey, W. K. (1992), “Estimating exposure effects by modeling the expectation of exposure conditional on confounders,” Biometrics, 48, 479-495.

[8] Robins JM, Rotnitzky A (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer” Statistica Sinica, 11(4):920-936.

[9] Tchetgen Tchetgen E, Robins J, Rotnitzy A (2010) On doubly robust estimation in a semi-parametric odds ratio model. Biometrika (2010) vol. 97(1), pages 171-180.

[10] Tchetgen Tchetgen E, Robins J. (2010a) The semiparametric case-only estimator. Biometrics (in Press), Published online March 10 2010.

[11] Tchetgen Tchetgen E . (2010a). Robust Discovery of Genetic Associations incorporating Gene-Environment Interaction and Independence. Epidemiology. In press.

[12] Vansteelandt S, VanderWeele TJ, Tchetgen Tchetgen E, Robins J.M., Multiply robust inference for statistical interactions (2008). *Journal of the American Statistical Association*, Vol. 103, No. 484, Theory and Methods

APPENDIX

PROOF OF THEOREM 1:

Let P denote the law of $(Y, \mathbf{A}, \mathbf{X})$. Suppose first that $q_2(\mathbf{X}, A_2)$ and $q_1(\mathbf{X}, A_1)$ are known functions, then model $\mathcal{A}$ is the semiparametric model

$$\operatorname{logit}\Pr(Y=1|\mathbf{A},\mathbf{X}) = m(\mathbf{X}, A; \beta^*)$$

where $m(\mathbf{X}, A; \beta^*) = q_3(A, \mathbf{X}; \beta^*) + q_2(\mathbf{X}, A_2) + q_1(\mathbf{X}, A_1) + h(\mathbf{X})$ with $h(\mathbf{X})$ an unknown function and the joint law of $(A, \mathbf{X})$ is unknown. Bickel et al (1993) showed that the orthocomplement to the nuisance tangent space for this model in the Hilbert space $L_2^0(P)$ (with covariance inner product) of functions in $L_2(P) = L_2$ with mean zero is given by

$$\Lambda_{nuis}^{SRLogit,\perp} = \{d(\mathbf{A},\mathbf{X})\}\,\epsilon(\beta^*) : E\left(d(\mathbf{A},\mathbf{X})\,\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{A},\mathbf{X})\,|\mathbf{X}\right) = 0\}$$

Consider now the original model $\mathcal{A}$ with $q_2(\mathbf{X}, A_2)$ and $q_1(\mathbf{X}, A_1)$ unrestricted. Consider a one-dimensional submodel $q_1(\mathbf{X}, A_1; t) = q_1(\mathbf{X}, A_1) + tk_1(\mathbf{X}, A_1)$. Then the score $S_t(\epsilon, A, \mathbf{X})$ for t at the truth $t = 0$ is of the form $\epsilon(\beta^*)\,k_1(\mathbf{X}, A_1)$ .Thus, $d(\mathbf{A},\mathbf{X})\,\epsilon(\beta^*) \in \Lambda_{nuis}^{SRLogit,\perp}$ must satisfy $E\left(d(\mathbf{A},\mathbf{X})\,\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{A},\mathbf{X})\,|A_1,\mathbf{X}\right)$. By symmetry, we conclude that the orthocomplement of the nuisance tangent space in model $\mathcal{A}$ is

$$\Lambda_{nuis}^{\perp} = \left\{\epsilon d(A,\mathbf{X})\,; E\{d(A,\mathbf{X})\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{A},\mathbf{X})\,|A_1,\mathbf{X}\} = E\{d(A,\mathbf{X})\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{A},\mathbf{X})\,|A_2,\mathbf{X}\} = 0\right\} \tag{A.1}$$

PROOF OF LEMMAS 1 & 2: Consider the Hilbert space $L_2\left(P^w_{\mathbf{A},\mathbf{X}}\right)$ of functions of $(\mathbf{A},\mathbf{X})$ under the tilted law of $[\mathbf{A}|X]$,$f^w(\mathbf{A}|\mathbf{X};\beta^*) = f(\mathbf{A}|\mathbf{X};\beta^*)\,\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{A},\mathbf{X})\left\{\int f(\mathbf{a}|\mathbf{X};\beta^*)\,\sigma^2_{Y|\mathbf{A},\mathbf{X}}(\mathbf{a},\mathbf{X})\,d\mu(\mathbf{a})\right\}^{-1}$

and the density of X as in model $\mathcal{A}$. Then we wish to find a characterization of functions in $L_2\left(P^w_{\mathbf{A},\mathbf{X}}\right)$ that satisfy

$$\mathcal{D} = \{d(A,\mathbf{X}) : E^w\{d(A,\mathbf{X})|A_1,\mathbf{X}\} = E^w\{d(A,\mathbf{X})|A_2,\mathbf{X}\} = 0\} \cap L_2\left(P^w_{\mathbf{A},\mathbf{X}}\right)$$

where E^w is the expectation under $f^w\left(\mathbf{A}|\mathbf{X};\beta^*\right)$. Lemma 2 is a corollary of Theorem 1 of Tchetgen Tchetgen et al (2010) who established that given an admissible density $h^\dagger\left(\mathbf{A}|\mathbf{X}\right) = g_1^\dagger\left(A_1|\mathbf{X}\right) g_2^\dagger\left(A_2|\mathbf{X}\right)$,

$$\mathcal{D} = \left\{ \ \{\mathbf{s}\left(\mathbf{A}\ ,\mathbf{X}\ \right) - \mathbf{s}^\dagger\left(\mathbf{A}\ ,\mathbf{X}\ \right)\}\frac{h^\dagger\left(\mathbf{A}\ |\mathbf{X}\ \right)}{f^w\left(\mathbf{A}\ |\mathbf{X}\ ;\beta^*\right)} : \mathbf{s}\left(\mathbf{A}\ ,\mathbf{X}\ \right)\right\} \cap L_2\left(P^w_{\mathbf{A},\mathbf{X}}\right)$$

Then, to get the result, it suffices to note that

$$\begin{aligned}
&\epsilon\left(\beta^*\right)\ \{\mathbf{s}\left(\mathbf{A}\ ,\mathbf{X}\ \right) - \mathbf{s}^\dagger\left(\mathbf{A}\ ,\mathbf{X}\ \right)\}\frac{h^\dagger\left(\mathbf{A}\ |\mathbf{X}\ \right)}{f^w\left(\mathbf{A}\ |\mathbf{X}\ ;\beta^*\right)} \\
&= \frac{\epsilon\left(\beta^*\right)}{\sigma^2_{Y|\mathbf{A},\mathbf{X}}\left(\mathbf{A},\mathbf{X}\right)}\ \{\mathbf{s}\left(\mathbf{A}\ ,\mathbf{X}\ \right) - \mathbf{s}^\dagger\left(\mathbf{A}\ ,\mathbf{X}\ \right)\}\frac{h^\dagger\left(\mathbf{A}\ |\mathbf{X}\ \right)}{f\left(\mathbf{A}|\mathbf{X};\beta^*\right)}\left\{\int f\left(\mathbf{a}|\mathbf{X};\beta^*\right)\sigma^2_{Y|\mathbf{A},\mathbf{X}}\left(\mathbf{a},\mathbf{X}\right)d\mu\left(\mathbf{a}\right)\right\} \\
&= \{\mathbf{s}_2\left(\mathbf{A}\ ,\mathbf{X}\ \right) - \mathbf{s}_2^\dagger\left(\mathbf{A}\ ,\mathbf{X}\ \right)\}\frac{h^\dagger\left(\mathbf{A}\ |\mathbf{X}\ \right)}{f\left(Y\ ,\mathbf{A}\ |\mathbf{X}\ ;\beta^*\right)}\left(-1\right)^{1-Y}
\end{aligned}$$

with $\mathbf{s}_2\left(\mathbf{A}_i,\mathbf{X}_i\right)=\left\{\int f\left(\mathbf{a}|\mathbf{X};\beta^*\right)\sigma^2_{Y|\mathbf{A},\mathbf{X}}\left(\mathbf{a},\mathbf{X}\right)d\mu\left(\mathbf{a}\right)\right\}\mathbf{s}_2\left(\mathbf{A},\mathbf{X}\right)$.

The equivalent representation of $\mathcal{D}$ provided in Section 4 of Tchetgen Tchetgen et al (2010) proves Lemma 1.

PROOF OF THEOREM 2

Lemma 1 implies that for any choice of $\mathbf{k}$, there exist an $\mathbf{s} = \mathbf{s}(\mathbf{k})$ such that $\mathbf{U}_i(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*); \mathbf{k}) = \mathbf{V}(\beta^*; \mathbf{s} = \mathbf{s}(\mathbf{k})) = \{\mathbf{s}(\mathbf{A}\ , \mathbf{X}\ ) - \mathbf{s}^\dagger(\mathbf{A}\ , \mathbf{X}\ )\}\frac{h^\dagger(\mathbf{A}\ |\mathbf{X}\ )}{f(Y\ , \mathbf{A}\ |\mathbf{X}\ ;\beta^*)}(-1)^{1-Y}$. We begin by showing that

$$E\{\mathbf{U}_i(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*); \mathbf{k})\ \} \tag{15}$$
$$= E\{\mathbf{V}(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*); \mathbf{s} = \mathbf{s}(\mathbf{k}))\} = 0$$

under model $\mathcal{B}$ where again we use $\tilde{}$ to denote probability limits. First, $\tilde{\gamma}(\beta^*) = \gamma^*$ under model $\mathcal{A} \cap \mathcal{M}_y$ because $\mathbf{H}_i(\beta^*, \tilde{\gamma}(\beta^*), \alpha)$ has mean zero for each α under this model. Equality (15) now follows because $E(\frac{(-1)^{1-Y_i}}{f(Y_i, \mathbf{A}_i|\mathbf{X}_i; \beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))}|\mathbf{A}, \mathbf{X}) = E(\frac{(-1)^{1-Y}}{f(Y|\mathbf{A}, \mathbf{X}; \beta^*, \gamma^*)}|\mathbf{A}, \mathbf{X})\frac{1}{f(\mathbf{A}_i|\mathbf{X}_i; \beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))} = 0$. Second, $(\tilde{\alpha}_0(\beta^*), \tilde{\alpha}_2(\beta^*), \gamma_2(\beta^*)) = (\alpha_0^*, \alpha_2^*, \gamma_2^*)$ under model $\mathcal{A} \cap \mathcal{M}_{ya2}$ (Tchetgen Tchetgen et al, 2010). Equality (15) now follows because

$$E\{\ \{\mathbf{s}(\mathbf{A}\ , \mathbf{X}\ ) - \mathbf{s}^\dagger(\mathbf{A}\ , \mathbf{X}\ )\}\frac{h^\dagger(\mathbf{A}\ |\mathbf{X}\ )}{f(Y\ , \mathbf{A}\ |\mathbf{X}\ ; \beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))}(-1)^{1-Y}\ \ |\ A_1, Y, \mathbf{X}\}$$
$$= E\{\ \{\mathbf{s}(\mathbf{A}\ , \mathbf{X}\ ) - \mathbf{s}^\dagger(\mathbf{A}\ , \mathbf{X}\ )\}\frac{h^\dagger(\mathbf{A}\ |\mathbf{X}\ )}{f(A_2|A_1, Y\ , \mathbf{X}\ ; \beta^*, \tilde{\alpha}(\beta^*))\, f(A_1 Y|\mathbf{X}; \beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))}(-1)^{1-Y}\ \ |\ A_1, Y, \mathbf{X}$$
$$= E^\dagger\{\ \{\mathbf{s}(\mathbf{A}\ , \mathbf{X}\ ) - \mathbf{s}^\dagger(\mathbf{A}\ , \mathbf{X}\ )\}\ (-1)^{1-Y}\ \ |\ A_1, Y, \mathbf{X}\}\frac{g_1^\dagger(A_1|\mathbf{X})}{f(A_1 Y|\mathbf{X}; \beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}(\beta^*))}$$
$$= 0$$

Third, equality (15) holds under model $\mathcal{A} \cap \mathcal{M}_{ya2}$ by symmetry.

Assuming that the regularity conditions of Theorem 1A in Robins, Mark and Newey (1992) hold for $U_i(\beta, \gamma, \alpha)$, $G_i(\beta, \gamma, \alpha)$ and $A_i(\alpha)$, it now follows by standard Taylor expansion arguments that

$$0 = n^{-1/2} \sum_{i=1}^{n} \mathbf{U}_i(\beta^*, \tilde{\theta}\,(\beta^*)\,; \mathbf{k}) + \Bigg[E \left\{ \frac{\partial}{\partial \beta} \mathbf{U}(\beta, \tilde{\theta}\,(\beta^*) \;\; ; \mathbf{k}) \;\right\}_{|\beta=\beta^*} - E \left\{ \frac{\partial}{\partial \theta} \mathbf{U}(\beta^*, \theta\;; \mathbf{k}) \right\}_{|\theta=\tilde{\theta}(\beta^*)}$$

$$\times E^{-1} \left\{ \frac{\partial}{\partial \theta} \mathbf{R}(\beta^*, \theta) \right\}_{|\theta=\tilde{\theta}(\beta^*)} E \left\{ \frac{\partial}{\partial \beta} \mathbf{R}(\beta, \tilde{\theta}\,(\beta^*)) \right\}_{|\beta=\beta^*} \Bigg] \sqrt{n}(\widehat{\beta} - \beta^*)$$

$$- E \left\{ \frac{\partial}{\partial \theta} \mathbf{U}(\beta^*, \theta \right\}_{|\theta=\tilde{\theta}(\beta^*)} E^{-1} \left\{ \frac{\partial}{\partial \theta} \mathbf{R}(\beta, \theta \right\}_{|\theta=\tilde{\theta}(\beta^*)} \mathbf{R}(\beta^*, \tilde{\theta}\,(\beta^*)) + o_p\,(1)$$

where $o_p(1)$ denotes a random variable converging to 0 in probability. When

$$\Bigg[E \left\{ \frac{\partial}{\partial \beta} \mathbf{U}(\beta, \tilde{\theta}\,(\beta^*) \;\; ; \mathbf{k}) \right\}_{|\beta=\beta^*} - E \left\{ \frac{\partial}{\partial \theta} \mathbf{U}(\beta^*, \theta\;; \mathbf{k}) \right\}_{|\theta=\tilde{\theta}(\beta^*)}$$

$$\times E^{-1} \left\{ \frac{\partial}{\partial \theta} \mathbf{R}(\beta^*, \theta) \right\}_{|\theta=\tilde{\theta}(\beta^*)} E \left\{ \frac{\partial}{\partial \beta} \mathbf{R}(\beta, \tilde{\theta}\,(\beta^*)) \right\}_{|\beta=\beta^*} \Bigg]$$

is nonsingular, it now follows that

$$\sqrt{n}(\widehat{\beta} - \beta^*) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} E^{-1} \left\{ \frac{\partial}{\partial \beta} \mathbf{U}^*(\beta, \tilde{\theta}\,(\beta^*)\,; \mathbf{k})|_{\beta=\beta^*} \right\} \mathbf{U}_i^*(\beta^*, \tilde{\theta}\,(\beta^*)\,; \mathbf{k}) + o_p(1) \qquad (16)$$

The asymptotic distribution of $\sqrt{n}(\widehat{\beta} - \beta^*)$ under model $\mathcal{B}$ follows from the previous equation by Slutsky's Theorem and the Central Limit Theorem. This proves part (i).

At the intersection model $\mathcal{A} \cap \mathcal{M}_y \cap \mathcal{M}_{ya_1} \cap \mathcal{M}_{ya_2}$,

$$E \left\{ \frac{\partial}{\partial \theta} \mathbf{U}(\beta, \theta; \mathbf{k}) \right\}_{|\theta=\tilde{\theta}(\beta^*)} = 0,$$

hence

$$\mathbf{U}^*(\beta^*, \tilde{\theta}\,(\beta^*)\,; \mathbf{k}) = \mathbf{U}(\beta^*, \tilde{\theta}\,(\beta^*)\,; \mathbf{k}).$$

It follows that the estimators $\widehat{\beta}(\mathbf{k}, \mathbf{G}_{(1)}, \mathbf{H}_{(1)})$ and $\widehat{\beta}(\mathbf{k}, \mathbf{G}_{(2)}, \mathbf{H}_{(2)})$ have the same influence functions at the intersection model $\mathcal{A} \cap \mathcal{M}_y \cap \mathcal{M}_{ya_1} \cap \mathcal{M}_{ya_2}$. This proves part (ii).

To prove part (iii) of the Theorem it suffices to show that the following equality holds:

$$E\left[\left\{\mathbf{S}_\beta\left(\beta^*\right) - \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k}_{opt})\right\} \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] = 0$$

for all scalar functions $\mathbf{k} \equiv \mathbf{k}\left(\mathbf{X}\right)$, where $\mathbf{S}_\beta\left(\beta^*\right) = Y \frac{\partial q_3(\mathbf{A},\mathbf{X};\beta)}{\partial \beta}|_{\beta=\beta^*} - E\left(Y \frac{\partial q_3(\mathbf{A},\mathbf{X};\beta)}{\partial \beta}|_{\beta=\beta^*}|\mathbf{X}\right)$ denotes the score of β^* in model $\mathcal{A}$. Since $E\left[E\left(Y \frac{\partial q_3(\mathbf{A},\mathbf{X};\beta)}{\partial \beta}|_{\beta=\beta^*}|\mathbf{X}\right) \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] = 0$ for all for all $\mathbf{k}$, we must check

$$E\left[\left\{Y \overline{\mathbf{q}}_3\left(A_2, \mathbf{X}\right) A_1 - \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k}_{opt})\right\} \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] = 0$$

where $\overline{\mathbf{q}}_3\left(A_2, \mathbf{X}\right) = \left(\frac{\partial q_3(z_1,\mathbf{X};\beta)}{\partial \beta}|_{\beta=\beta^*}, ..., \frac{\partial q_3(z_J,\mathbf{X};\beta)}{\partial \beta}|_{\beta=\beta^*}\right)$

We note that

$$\begin{aligned}
&E\left[Y \overline{\mathbf{q}}_3\left(A_2, \mathbf{X}\right) A_1 \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] \\
&= E\left[\epsilon\left(\beta^*\right) \overline{\mathbf{q}}_3\left(A_2, \mathbf{X}\right) A_1 \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] \\
&= E\left[\epsilon\left(\beta^*\right) \overline{\mathbf{q}}_3\left(A_2, \mathbf{X}\right) \Delta \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right] \\
&= E\left[\mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k}_{opt}) \mathbf{U}(\beta^*, \tilde{\theta}\left(\beta^*\right); \mathbf{k})\right]
\end{aligned}$$

proving the result.

Table 1. Simulation results

		Logistic Regression	Multiply Robust
		β	β
$\mathcal{M}_y$ correct	bias	0.005	0.009
	variance	0.092	0.106
$\mathcal{M}_{ya_1}$ correct	bias	0.113	-0.002
	variance	0.093	0.104
$\mathcal{M}_{ya_2}$ correct	bias	0.119	0.030
	variance	0.094	0.113
All models wrong	bias	0.151	0.119
	variance	0.100	0.108