

Estimation of controlled direct effects on a dichotomous outcome using logistic structural direct effect models

BY S. VANSTEEELANDT

*Department of Applied Mathematics and Computer Sciences, Ghent University, 281 (S9)
Krijgslaan, 9000 Ghent, Belgium
stijn.vansteelandt@ugent.be*

SUMMARY

We consider the problem of assessing whether an exposure affects a dichotomous outcome other than by modifying a given mediator. The standard approach, logistic regression adjusting for both exposure and mediator, is known to be biased in the presence of confounders for the mediator-outcome relationship. Because additional regression adjustment for such confounders is only justified when they are not affected by the exposure, inverse probability weighting has been advocated, but is not ideally tailored to mediators that are continuous or have strong measured predictors. We overcome this limitation by developing inference for a novel class of causal models which are closely related to Robins' logistic structural direct effect models, but do not inherit their difficulties of estimation. We study identification and efficient estimation under the assumption that all confounders for the exposure-outcome and mediator-outcome relationships have been measured, and find adequate performance in simulation studies. We discuss extensions to case-control studies and relevant implications for the generic problem of adjustment for time-varying confounding.

Some key words: Case-control study; Direct effect; G-estimation; Instability; Logistic structural mean model; Mediation; Pathway; Time-varying confounding.

1. INTRODUCTION

The statistical problem of inferring the extent to which the effect of one variable on another is mediated by some possible intermediate variable, has been the subject of considerable research (MacKinnon, 2008), with an approach based on regression analysis (Baron & Kenny, 1986) now being routinely utilized. Over the past decade, this problem has gained renewed attention within the causal inference literature, mainly in epidemiology and biostatistics. This has lead to better phrased definitions of direct and indirect effects, to an improved understanding of the conditions under which these can be identified (Robins & Greenland, 1992; Pearl, 2001; Rubin, 2004; VanderWeele & Vansteelandt, 2009) and to novel estimation strategies that are valid under weaker causal assumptions and/or less restrictive modelling assumptions (Robins, 1999b; Petersen et al., 2006; Joffe et al., 2007; Goetgeluk et al., 2008; van der Laan & Petersen, 2008; Emsley et al., 2009; Gallop et al., 2009; VanderWeele, 2009; Vansteelandt, 2009; Imai et al., 2010; VanderWeele & Vansteelandt, 2009, 2010).

In this article, we focus on the estimation of controlled direct effects (Robins & Greenland, 1992; Pearl, 2001). To define these, let $Y(x, m)$ denote the potential outcome that would, possibly contrary to the facts, have been observed for given subject if the exposure were set to x and the mediator to m . Then a controlled direct effect is a comparison of potential outcomes corre-

sponding to different exposure levels x , but the same mediator level m . This can for instance be expressed in terms of an odds ratio

$$\frac{\text{odds}\{Y(x, m) = 1\}}{\text{odds}\{Y(0, m) = 1\}}, \quad (1)$$

where for any 2 random variables A and B , $\text{odds}(A = 1 | B) = \text{pr}(A = 1 | B) / \text{pr}(A = 0 | B)$. Identification of controlled direct effects requires separate no unmeasured confounders assumptions regarding both the exposure-outcome and mediator-outcome relationships (VanderWeele & Vansteelandt, 2009). In particular we will assume throughout that X is a randomly assigned exposure and that controlling for a given, measured set of confounders L in addition to the exposure X , is sufficient for estimating the causal effect of M on Y . These assumptions hold under the causal diagram (Pearl, 2000) of Figure 1 and can be formalized as

$$Y(x, m) \perp\!\!\!\perp X \quad (2)$$

$$Y(x, m) \perp\!\!\!\perp M | L, X = x \quad (3)$$

for all x, m ; see Appendix 3 for relaxations of assumption (2) to allow for confounding by measured factors.

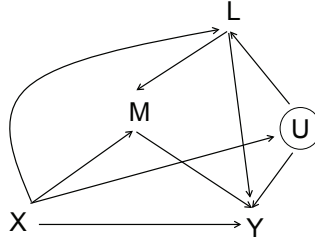


Fig. 1: Causal diagram with measured variables X, L, M and Y , and with U an unmeasured confounder of the L - Y relationship.

Estimation of controlled direct effects under assumptions (2) and (3) has been the subject of considerable recent research; see Vansteelandt (2009) for an overview. Much attention has been given to settings where the confounder L is potentially affected by the exposure, in which case we refer to L as an intermediate confounder. Such settings are common, especially when the mediator comes much later in time than the exposure itself (VanderWeele & Vansteelandt, 2009). They call for specialized estimation methods since standard regression adjustment for post-exposure measurements, such as L , may induce a collider-stratification bias (Cole & Hernán, 2002). Robins (1999b) and VanderWeele (2009) avoid regression adjustment for L . They instead control for confounding due to L by weighting each subject's data by the reciprocal of the conditional mediator density $f(M | X, L)$. In particular, a weighted logistic regression of outcome on exposure and mediator yields a consistent estimator of the controlled direct effect (1) (VanderWeele, 2009). Vansteelandt (2009) examined this approach in simulation studies with a continuous mediator and outcome, considering the wide interest in this case (MacKinnon, 2008; Imai et al., 2010). He diagnosed poor performance of the inverse probability weighted estimates, sometimes having substantially greater finite-sample bias than the corresponding unweighted estimates which ignore confounding altogether. Goetgeluk et al. (2008) attained considerable improvements under these conditions via more efficient estimators. However, instability remained

because minor estimation errors in the tail of a continuous density $f(M | X, L)$ can have a major influence on the weights. For risk differences and risk ratios, an alternative is G-estimation for additive and multiplicative structural nested direct effect models (Robins, 1999a). This involves first regressing the outcome on the mediator, exposure and confounders to infer the mediator's effect on the outcome, and next assessing the residual association with the exposure after removing that effect from the outcome. Joffe & Greene (2009), Vansteelandt et al. (2009) and Vansteelandt (2009) study this in the case of a continuous outcome and mediator that lend themselves to linear modeling, and find a much improved performance. However, additive and multiplicative models for a dichotomous outcome are prone to misspecification and a similar development for odds ratios has turned out to be impossible so far (Robins, 1999a). In this article, we aim to overcome the aforementioned limitations via a novel class of causal models for controlled direct effect odds ratios which are closely related to Robins' logistic direct effect structural mean models, but avoid the need for repeated averaging that prohibits estimation under these models.

2. MODEL AND ESTIMATION

2.1. Logistic direct effect structural mean models

We consider direct effect models that parameterize the controlled direct effect odds ratio conditional on exposure and intermediate confounders:

$$\frac{\text{odds}\{Y(x, 0) = 1 \mid X = x, L(x)\}}{\text{odds}\{Y(0, 0) = 1 \mid X = x, L(x)\}} = \exp \{\gamma(x; \beta^*)\}, \quad (4)$$

where $L(x)$ refers to the counterfactual value of L at exposure level x , where $\gamma(X; \beta)$ is an arbitrary function of X , smooth in β and satisfying $\gamma(0; \beta) = 0$, and where β^* is an unknown, finite-dimensional parameter; e.g., $\gamma(X; \beta) = \beta X$. Model (4) is defined like a logistic structural direct effect model (Robins, 1999a,b), except that it conditions additionally on the intermediate confounder $L(x)$. This additional conditioning does not compromise the validity of (4) as a direct causal effect parameter. This can be seen since it involves comparing the same subset of the population, $\{X = x, L(x)\}$, twice under different interventions. In particular, the counterfactual $L(x)$ can be viewed as a baseline variable so that (4) does not involve conditioning on a post-exposure variable (Joffe et al., 2007). Consider for instance the HIV-prevention trial analyzed in Rosenblum et al. (2009). Let $X = 1$ denote randomized assignment to the use of diaphragm and microbicides at all sex acts versus $X = 0$ encoding no intervention, M measure the percentage of sexual acts protected with the male condom, L measure the number of male sexual partners and $Y = 1$ denote HIV infection after 24 months of follow-up. Then

$$\frac{\text{odds}\{Y(1, 0) = 1 \mid X = 1, L(1) = 1\}}{\text{odds}\{Y(0, 0) = 1 \mid X = 1, L(1) = 1\}}$$

expresses how much, if no male condoms were available, the odds of HIV infection after 24 months of follow-up would change for women with a single male sexual partner on the intervention arm, if they were assigned to no intervention. This change may be partly attributable to intervention-induced changes in the number of male sexual partners or other factors than male condom use, as expressed through the pathways $X \rightarrow L \rightarrow Y$ and $X \rightarrow U \rightarrow Y$ in Figure 1.

The decision to measure direct effects within levels of the confounder L is made here for the same reason that it is commonly made in the simpler setting where L is assumed not to be affected by X , namely to enable a feasible estimation approach. Under this assumption that L is not affected by X , direct effects are usually estimated via ordinary regression of outcome Y on

exposure X , involving mediator M and confounders L :

$$\text{logit}E(Y \mid X, L, M) = \mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) + \mu_m(X, L, M; \eta^*), \quad (5)$$

where $\text{logit}(x) = \log\{x/(1-x)\}$. Here, the main effects $\mu_x(X; \eta)$ of X , $\mu_l(X, L; \eta)$ of L and $\mu_m(X, L, M; \eta)$ of M are arbitrary known functions of (X, L, M) , smooth in η and satisfying $\mu_l(X, 0; \eta) = \mu_m(X, L, 0; \eta) = 0$, and η^* is an unknown finite-dimensional parameter. Suppose furthermore that the usual consistency assumptions hold that $Y(x, m)$ equals the observed outcome Y amongst those with $X = x$ and $M = m$ and that $L(x)$ equals the observed confounder L amongst those with $X = x$. Then it follows from (3) that $E(Y \mid X = x, L = l, M = 0) = E\{Y(x, 0) \mid X = x, L(x) = l\}$. When L is not affected by X so that (2) can be strengthened to $Y(0, 0) \perp\!\!\!\perp X \mid L$, then it further follows that $E(Y \mid X = 0, L = l, M = 0) = E\{Y(0, 0) \mid X = x, L(x) = l\}$ for all x . Upon combining these results it is seen that $\exp\{\mu_x(X; \eta^*) - \mu_x(0; \eta^*) + \mu_l(X, L; \eta^*) - \mu_l(0, L; \eta^*)\}$ in model (5) captures the direct effect (4) when L is not affected by X . Our focus on the same direct effect estimand as a standard regression analysis in cases where the latter is valid, will enhance the comparability of the proposed approach with routine practice. In particular, when L is affected by X so that standard regression adjustment fails, a comparison of $\mu_x(X; \eta^*) - \mu_x(0; \eta^*) + \mu_l(X, L; \eta^*) - \mu_l(0, L; \eta^*)$ in model (5) and $\gamma(x; \beta^*)$ in model (4) can be insightful in the sense of quantifying the degree of bias due to collider-stratification along the undirected path $X \rightarrow L \leftarrow U \rightarrow Y$ in Figure 1 and/or due to blocking an indirect effect mediated by L . For instance, suppose that $\gamma(X; \beta) = \beta X$, $\mu_x(X; \eta) = \eta_0 + \eta_x X$ and $\mu_l(X, L; \eta) = \eta_l L$. Then $\exp(\eta_x - \beta)$ can be viewed as a measure of bias due to inappropriate adjustment for intermediate confounding because it expresses how much the direct effect odds ratio $\exp(\beta)$ differs from the corresponding association measure $\exp(\eta_x)$. Similar measures are popular amongst epidemiologists to express the strength of confounding (Miettinen, 1972), but are not directly obtainable via inverse probability weighted estimators (VanderWeele, 2009). The latter target the unconditional direct effects odds ratio (1) which tends to be different in magnitude in view of noncollapsibility of the odds ratio.

A limitation of the direct effects odds ratio (4) is that it cannot be used for making treatment decisions when L is not known at the time of administering the treatment. Similar limitations hold for various other direct effect measures (Rubin, 2004; Joffe et al., 2007). We do not consider this to be a major drawback because the subset $\{X = x, L(x)\}$ is identifiable and because the primary purpose of measuring direct effects is typically much rather to develop a better understanding of the causal mechanism. Furthermore, we show in §2.3 how estimates of β^* can be used as plug-in estimates when drawing inference for the marginal direct effects odds ratio

$$\frac{\text{odds}\{Y(X, 0) = 1\}}{\text{odds}\{Y(0, 0) = 1\}}. \quad (6)$$

The latter is of greater relevance for policy making as it expresses the effect of setting each subject's exposure to zero, while controlling the mediator at zero.

2.2. Identification

We use $\mathcal{A}$ to denote the model defined by the no unmeasured confounders assumptions (2) and (3), the aforementioned consistency assumptions, the structural model (4) and the association model (5). Model (5) and the no unmeasured confounders assumption (3) imply

$$E\{Y(X, 0) \mid X, L\} = E\{Y(X, 0) \mid X, L, M = 0\} = \text{expit}\{\mu_x(X; \eta^*) + \mu_l(X, L; \eta^*)\}$$

where $\text{expit}(x) = \exp(x)/\{1 + \exp(x)\}$, and the structural model (4) subsequently implies

$$E\{Y(0, 0) \mid X, L\} = \text{expit}\{\mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*)\}.$$

It follows that $\gamma(X; \beta^*)$ can be identified from the identity $E\{Y(0,0) | X\} = E\{Y(0,0)\}$, which is implied by the no unmeasured confounders assumption (2). Proposition 1 below, whose proof is given in Appendix 1, suggests that this is the only restriction on the observed data which is implied by model (4).

PROPOSITION 1. *Model $\mathcal{A}$ imposes the same restrictions on the observed data as model $\mathcal{B}$, which is defined by (5) and the mean independence restriction*

$$\begin{aligned} E[\text{expit}\{\mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*)\} | X] \\ = E[\text{expit}\{\mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*)\}], \end{aligned} \quad (7)$$

provided that, for the law $f(L | X)$, there exists a value β^* satisfying (7) for all X .

REMARK 1. *Because equation (7) implies one restriction for each unknown $\gamma(x; \beta^*)$ for each $x \neq 0$ in the support of X , a dependence of the controlled direct effect in (4) on $L(x)$ is not nonparametrically identified. We do not consider this limitation to be prohibiting because, as usual, ignoring interactions leads to approximately ‘averaged’ effects. This is confirmed through simulation experiments in §4, as well as theoretically in Appendix 2 where we consider linear structural mean models defined like (4), but with the logit-link replaced with the identity link. We then find that, while $\gamma(x; \hat{\beta})$, where $\hat{\beta}$ is an estimator of β^* as derived in the §2.3, may not be a consistent estimator of $E\{Y(x, m) | X = x, L(x)\} - E\{Y(0, m) | X = x, L(x)\}$, it is a consistent estimator of $E\{Y(x, m) | X = x\} - E\{Y(0, m) | X = x\}$ despite interactions with $L(x)$ being ignored. The use of, often stronger, untestable no-interaction assumptions is not uncommon in causal inference, e.g. in instrumental variable analysis (Hernán & Robins, 2006), in the estimation of natural direct effects (Robins & Greenland, 1992; Petersen et al., 2006) and for handling more general identification problems (Vansteelandt & Goetghebeur, 2004; Ten Have et al., 2007).*

2.3. Estimation

An estimate $\hat{\eta}$ of η^* can be straightforwardly obtained by fitting model (5) via standard logistic regression methods. Following Proposition 1, estimation of β^* can happen to ensure that (7) holds; for instance, by solving the following estimating equation:

$$0 = \sum_{i=1}^n \{X_i - E(X_i)\} \text{expit}\{\mu_x(X_i; \hat{\eta}) + \mu_l(X_i, L_i; \hat{\eta}) - \gamma(X_i; \beta)\},$$

for β . More generally, we demonstrate in Appendix 3 that a semi-parametric efficient estimator of $\lambda^* \equiv (\beta^{*T}, \eta^{*T})^T$ in model $\mathcal{A}$ can be obtained by solving an estimating equation of the form

$$\begin{aligned} 0 = \sum_{i=1}^n U_i(\beta, \eta) = \sum_{i=1}^n d(M_i, L_i, X_i) [Y_i - \text{expit}\{m(M_i, L_i, X_i; \eta)\}] \\ + [e(X_i) - E\{e(X_i)\}] \text{expit}\{\mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) - \gamma(X_i; \beta)\}, \end{aligned} \quad (8)$$

jointly for β and η , with $m(M, L, X; \eta) \equiv \mu_x(X; \eta) + \mu_l(X, L; \eta) + \mu_m(M, L, X; \eta)$, $d(M, L, X) = \left(0_p^T, m_\eta^T(M, L, X; \eta)\right)^T$, with 0_p a $p \times 1$ vector of zeroes, p the dimension of β , and for any function $\phi(\cdot)$ of a parameter α , $\phi_\alpha(\cdot)$ denoting the first-order derivative $\partial\phi(\cdot)/\partial\alpha$. Further, with $\text{expit}'(x) = \text{expit}(x) \{1 - \text{expit}(x)\}$,

$$e(X) = \sigma^{-2}(X) \left[f(X) - E\left\{ \sigma^{-2}(X) f(X) \right\} / E\left\{ \sigma^{-2}(X) \right\} \right]$$

$$f(X) = E \left[\text{expit}' \{ \mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*) \} \left(\begin{matrix} \gamma_\beta(X; \beta^*) \\ \mu_{x,\eta}(X; \eta^*) + \mu_{l,\eta}(X, L; \eta^*) \end{matrix} \right) \mid X \right]$$

$$\sigma^2(X) = \text{var} [\text{expit} \{ \mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*) \} \mid X].$$

Estimating equation (8) is unbiased, and thus gives consistent estimators of β^* and η^* , for each choice of functions $d(M_i, L_i, X_i)$ and $e(X_i)$ provided standard regularity conditions hold. We discuss two simple choices of $e(X)$ that lead to estimators for β^* with attractive properties. The choice

$$e(X) = \begin{pmatrix} E [\text{expit}' \{ \mu_x(X; \eta^*) + \mu_l(X, L; \eta^*) - \gamma(X; \beta^*) \} \gamma_\beta(X; \beta^*) \mid X] \\ 0_q \end{pmatrix}, \quad (9)$$

where q is the dimension of η , allows for fitting the association model using off-the-shelf software for logistic regression and thus gives computationally simpler estimators, though with reasonable efficiency. It can be shown that the choice

$$e(X) = \begin{pmatrix} \gamma_\beta(X; \beta^*) \\ \mu_{x,\eta}(X; \eta^*) + E \{ \mu_{l,\eta}(X, L; \eta^*) \mid X \} \end{pmatrix} \quad (10)$$

gives locally robust estimators of β^* which remain consistent under model misspecification in the absence of a total effect, provided that $\gamma(X; \beta)$ is linear in β and nested within $\mu_x(X; \eta)$ and provided that, as usual, $\mu_x(X; \eta)$ carries a constant term for the intercept and the parameters indexing $\mu_x(X; \eta)$ and $\mu_l(X, L; \eta)$ are variation independent. Other locally robust estimators are given in Appendix 3. The asymptotic variance of an estimator $\hat{\lambda} = (\hat{\beta}^T, \hat{\eta}^T)^T$ for λ^* obtained by solving (8) follows by the sandwich formula $n^{-1} E \{ \partial U_i(\lambda^*) / \partial \lambda \}^{-1} \text{var} \{ U_i(\lambda^*) \} E \{ \partial U_i(\lambda^*) / \partial \lambda \}^{-1, T}$.

Given consistent estimators $\hat{\beta}$ of β^* and $\hat{\eta}$ of η^* , a consistent estimator of the marginal direct effects odds ratio (6) is given by $\hat{\theta} = \hat{p}(1 - \hat{p}_0) / \{ \hat{p}_0(1 - \hat{p}) \}$ for

$$\hat{p} = n^{-1} \sum_{i=1}^n \text{expit} \{ \mu_x(X_i; \hat{\eta}) + \mu_l(X_i, L_i; \hat{\eta}) \}$$

$$\hat{p}_0 = n^{-1} \sum_{i=1}^n \text{expit} \{ \mu_x(X_i; \hat{\eta}) + \mu_l(X_i, L_i; \hat{\eta}) - \gamma(X_i; \hat{\beta}) \}.$$

It can be shown that the asymptotic variance of $\log(\hat{\theta})$ can be estimated as 1 over n times the empirical variance of

$$\begin{aligned} & \frac{1}{p(1-p)} (\text{expit} \{ \mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) \} - p \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n E \left[\frac{\partial}{\partial \lambda} \text{expit} \{ \mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) \} \right] E \left\{ \frac{\partial U_i(\lambda)}{\partial \lambda} \right\}^{-1} U_i(\lambda) \Bigg) \\ & - \frac{1}{p_0(1-p_0)} (\text{expit} \{ \mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) - \gamma(X_i; \beta) \} - p_0 \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n E \left[\frac{\partial}{\partial \lambda} \text{expit} \{ \mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) - \gamma(X_i; \beta) \} \right] E \left\{ \frac{\partial U_i(\lambda)}{\partial \lambda} \right\}^{-1} U_i(\lambda) \Bigg), \end{aligned}$$

where p, p_0, η, β and λ can be replaced with $\hat{p}, \hat{p}_0, \hat{\eta}, \hat{\beta}$ and $\hat{\lambda}$, respectively.

2.4. Congenial models

As implicitly suggested in Proposition 1, for some η^* and laws $f(L | X)$, there may exist no value of β^* for which (7) holds for all X , thus contradicting assumption (2). When for any η^* and law $f(L | X)$, no such solution β^* to (7) exists, then model $\mathcal{A}$ is not well-defined. We suggest two ways to remedy this. The first overcomes this potential uncongeniality problem by leaving $\mu_{x0}(X) \equiv \mu_x(X; \eta^*) - \gamma(X; \beta^*)$ an unspecified function of X and estimating it so as to make (7) hold for all X , using the following iterative strategy. For given $\theta \in]0, 1[$ and given η indexing $\mu_l(X, L; \eta)$, find an estimate $\hat{\mu}_{x0}(X; \theta)$ of $\mu_{x0}(X)$ by solving the integral equation

$$\int \text{expit} \{ \mu_{x0}(X) + \mu_l(X, L; \eta) \} f(L | X) dL = \theta. \quad (11)$$

Next, given $\hat{\mu}_{x0}(X; \theta)$, estimate β, η and θ by maximizing the parametric likelihood

$$\prod_{i=1}^n \text{pr}(Y_i | X_i, L_i, M_i; \beta, \eta, \theta),$$

where

$$\text{pr}(Y | X, L, M; \beta, \eta, \theta) = \text{expit} \{ \hat{\mu}_{x0}(X; \theta) + \gamma(X; \beta) + \mu_l(X, L; \eta) + \mu_m(X, L, M; \eta) \}.$$

Upon iterating this process until convergence, a consistent estimator of β^* is obtained. This strategy is closely related to an approach advocated by Robins & Rotnitzky (2004) for estimating treatment effects in randomized trials with noncompliance and a dichotomous outcome. We will not pursue this strategy here because it requires modeling the joint law $f(L | X)$, which can be difficult when L is high-dimensional with both discrete and continuous components, and because it requires repeatedly solving integral equations of the form (11) for each observed value of X , which is computationally cumbersome.

We therefore recommend the strategy of §2.3, which is closely related to a recommendation by Vansteelandt & Goetghebeur (2003) in the context of instrumental variable estimation. This strategy is more practical because it avoids the need for solving integral equations and does not require modeling the distribution of the intermediate confounders conditional on X . When there is concern about uncongeniality, this can be remedied by relaxing the restrictions on the association model. Specifically, the model restrictions imposed by the first strategy are those of model $\mathcal{A}^*$, which is defined like model $\mathcal{A}$, but with the association model (5) redefined to be

$$\text{logit} E(Y | X, L, M) = \mu_x(X) + \mu_l(X, L; \eta^*) + \mu_m(X, L, M; \eta^*), \quad (12)$$

where $\mu_x(X)$ is an unknown function, and $\mu_l(X, L; \eta)$ and $\mu_m(X, L, M; \eta)$ are defined as before. This redefinition suggests a second strategy whereby we leave the main exposure effect in the association model unspecified to guarantee the existence of a value of β^* such that (7) holds for all X . For a categorical exposure, standard logistic regression software can be used for fitting this association model. For a continuous exposure X , off-the-shelf software for generalized additive models can be used.

REMARK 2. *The integral equation (7) can have a solution even when parametric restrictions are imposed on the main effect of X in the association model. For instance, when $\mu_l(X, L; \eta) = 0$, then L is not an intermediate confounder and hence, not surprisingly, (7) holds upon choosing $\gamma(X; \beta)$ equal to $\mu_x(X; \eta) - \mu_x(0; \eta)$. This suggests that $\mu_x(X; \eta)$ should at least have the complexity of $\gamma(X; \beta)$ in practice. Further, when $\mu_l(X, L; \eta) = 0$ then $\sigma^2(X) = 0$. It follows from the previous that the choice of $e(X)$ does not affect the efficiency of estimators of λ^* in that case, so we choose to set $e(X) = 1$ when $\sigma^2(X) = 0$.*

3. EXTENSIONS

3.1. Case-control studies

We show that a minor modification of the estimation principle of §2.3 yields approximately consistent estimators of (4) under a rare disease assumption. Let Y_i indicate 1 for cases and 0 for controls. Then modifying estimating equation (8) to

$$0 = \sum_{i=1}^n d(M_i, L_i, X_i) [Y_i - \text{expit} \{m(M_i, L_i, X_i; \eta)\}] + [e(X_i) - E\{e(X_i) \mid Y_i = 0\}] (1 - Y_i) \\ \times \text{expit} \{\mu_x(X_i; \eta) + \mu_l(X_i, L_i; \eta) - \gamma(X_i; \beta)\} \quad (13)$$

yields the desired result for the following reasons. First, under the rare disease assumption the controls form a fairly representative sample from the population, which explains why the second term in (13) is now evaluated in the controls. Second, it follows from Prentice & Pyke (1979) that the first term in (13), with $m(M_i, L_i, X_i; \eta)$ set to $m(M_i, L_i, X_i; \eta^*)$ apart from an additive constant, is a mean zero estimating function. Finally, while $\mu_x(X_i; \eta^*)$ is thus only estimable up to an additive constant, the second term in (13) is relatively insensitive to additive constants. This can be understood from the following argument: $E\{Y_i(0, 0) \mid X_i, L_i\} = \text{expit} \{\mu_x(X_i; \eta^*) + \mu_l(X_i, L_i; \eta^*) - \gamma(X_i; \beta^*)\}$ is mean independent of X_i under model $\mathcal{A}$. If the independence were to hold in distribution, then we would also have that $\text{logit} E\{Y_i(0, 0) \mid X_i, L_i\} = \mu_x(X_i; \eta^*) + \mu_l(X_i, L_i; \eta^*) - \gamma(X_i; \beta^*)$ is independent of X_i , suggesting that the second term in (13) has approximately mean zero at $\beta = \beta^*$, even when $\mu_x(X_i; \eta^*)$ is misspecified by an additive constant. In §4, we will study the accuracy of this approximation via simulation experiments. See Vansteelandt (2009) for approximate controlled direct effect relative risk estimators under a rare disease assumption.

3.2. Time-varying exposures

We demonstrate that the proposed methods extend relatively straightforwardly to enable estimation of the causal effect of a possibly continuous time-varying exposure on a dichotomous outcome in the presence of measured time-varying confounders (Robins et al., 2000). Let X_t and L_t be the exposure of interest and the set of measured confounders, respectively, at time $t = 1, \dots, T$. Suppose that X_t is possibly influenced by $\bar{L}_t \equiv (L_1, \dots, L_t)$ and that L_t is possibly influenced by $\bar{X}_{t-1} \equiv (X_1, \dots, X_{t-1})$. Let $Y(\bar{x}_T)$ and $\bar{L}_T(\bar{x}_T)$ denote the dichotomous potential outcome and confounder history at the end-of-study time upon setting the exposure history to $\bar{x}_T = (x_1, \dots, x_T)$. Throughout, we make the no unmeasured confounders assumption that $Y(\bar{x}_T) \perp\!\!\!\perp \bar{X}_t \mid \bar{X}_{t-1} = \bar{x}_{t-1}, \bar{L}_t$ for all $t = 1, \dots, T$ and all feasible exposure histories $\bar{x}_T$ (Robins, 1999b). Consider the following logistic structural model

$$\text{logit} E \{Y(\bar{x}_t, 0) \mid \bar{X}_T = \bar{x}_T, \bar{L}_T(\bar{x}_T)\} - \text{logit} E \{Y(\bar{x}_{t-1}, 0) \mid \bar{X}_T = \bar{x}_T, \bar{L}_T(\bar{x}_T)\} = \psi_t^* x_t$$

for $t = 1, \dots, T$. This model is closely linked to the logistic structural nested model (Robins, 1999b), except that, at each time t , it conditions on the exposure and confounder history up to the end-of-study time, rather than up to time t .

Using similar arguments as before, the unknown causal effect parameters ψ_t^* can be estimated by first fitting a standard association model

$$\text{logit} E \{Y(\bar{X}_T) \mid \bar{X}_T, \bar{L}_T\} = m(\bar{X}_T, \bar{L}_T; \eta^*),$$

where $m(\bar{X}_T, \bar{L}_T; \eta)$ is a known function, smooth in η , and η^* is an unknown finite-dimensional parameter. Next, we obtain that $\text{logit} E \{Y(\bar{X}_{t-1}, 0) \mid \bar{X}_T, \bar{L}_T\} =$

$m(\bar{X}_T, \bar{L}_T; \eta^*) - \sum_{s=t}^T \psi_s^* X_s$. This, together with the no unmeasured confounders assumption, implies that $\psi_t^*, t = 1, \dots, T$ can be estimated by recursively solving

$$0 = \sum_{i=1}^n \left[d_t(\bar{X}_{it}, \bar{L}_{it}) - E \left\{ d_t(\bar{X}_{it}, \bar{L}_{it}) \mid \bar{X}_{i,t-1}, \bar{L}_{it} \right\} \right] \expit \left\{ m(\bar{X}_{iT}, \bar{L}_{iT}; \eta^*) - \sum_{s=t}^T \psi_s^* X_{is} \right\},$$

where $d_t(\bar{X}_t, \bar{L}_t)$ is an arbitrary function of $(t, \bar{X}_t, \bar{L}_t)$, taking the dimension of ψ_t . While the considered logistic structural models are no longer nonparametrically identifiable for similar reasons as discussed in Remark 1, the main exposure effects are identifiable, along with interactions with the history of measured exposures and confounders. As with model (4), the suggested approach has the attractive feature of reducing to standard regression adjustment for time-varying confounders when the latter are not intermediate.

4. SIMULATION STUDY

In a first simulation experiment, we compared the validity and efficiency of the different estimators. We chose not to generate data conditional on an unmeasured variate U , as suggested in Figure 1, since a nonlinear model for $E(Y|X, M, L, U)$ does not generally imply a simple model for $E(Y|X, M, L)$. Instead, we generated 1000 measurements (X, L, M, Y) with X a normal variate with mean and variance 1, L and M normally distributed variates with mean $-1 + \alpha X$ and $-2 + 3X - 2L$, respectively, with variance 1 and α taking the values 0 and 1. We assumed that $E\{Y(0, 0)|X, L, M\} = \expit[\lambda + \eta_l\{L - E(L|X)\}]$, which satisfies (11). Its dependence on X when $\eta_l \neq 0$ reflects collider-stratification bias resulting from conditioning on the intermediate confounder L . Further, we assumed that the direct effects odds ratio given in (4) equals $\exp(\beta x + \eta_{xl}x[L(x) - E\{L(x)|X = x\}])$ and that $\text{odds}\{Y(x, m) = 1 \mid X = x, L\} / \text{odds}\{Y(x, 0) = 1 \mid X = x, L\} = \exp(-0.5m)$, implying that Y is a Bernoulli variate with mean $\expit[\lambda + \beta X + (\eta_l + \eta_{xl}X)\{L - E(L|X)\} - 0.5M]$.

Table 1: Validity and efficiency of competing estimators of the controlled and marginal direct effect log odds ratio in random population samples.

$\alpha, \eta_l, \eta_{xl}$	LR-M			LR-ML			LD-C			LD-M		
	Bias	SE	Cov	Bias	SE	Cov	Bias	SE	Cov	Bias	SE	Cov
1,1,0	3.84	1.0	3.0	-10.0	3.1	10.9	0.04	1.4	94.7	0.11	1.4	94.4
0,1,0	11.7	1.9	0.0	-0.07	3.4	95.6	-0.07	3.5	95.3	0.20	3.2	92.3
1,0,0	-0.03	0.9	94.8	0.02	2.9	94.4	-0.01	1.2	94.9	-0.01	1.4	93.9
1,1,0.1	4.66	1.1	0.1	-10.5	3.2	8.1	0.56	1.4	92.4	0.61	1.4	94.0
1,1,-0.1	2.99	1.0	15.5	-9.52	3.1	12.9	-0.47	1.3	93.2	-0.40	1.4	91.0
1,1,0.25	5.73	1.1	0.0	-11.1	3.3	6.5	1.24	1.5	84.4	1.27	1.1	87.9

Empirical bias (multiplied by 10), empirical standard error (multiplied by 10) and coverage (in %) of 95% confidence intervals for the standard logistic regression estimators, adjusting for M (LR-M) and (M, L) (LR-ML), and the proposed estimators of the conditional (LD-C) and marginal (LD-M) controlled direct effect log odds ratio.

We let λ and β take the values -1.5 and 0.25, η_l be 0 or 1 and η_{xl} take the values -0.1, 0.01 and 0.25, while setting η_{xl} equal to zero in the model fitting procedure to acknowledge the lack of identifiability of X - L interactions in the structural model. We next evaluated the standard logistic regression coefficient of X , adjusting for M only, and for both M and L . In addition, we evaluated the proposed estimator of the conditional controlled direct effect log odds ratio obtained by

solving (8) with $e(X) = (X, 0_q^T)^T$ and the proposed estimator of the marginal direct effects log odds ratio; see (6). We do not display results for the computationally convenient choice (9) and for the locally robust choice (10) because these were virtually identical. We also do not report results for the semi-parametric efficient estimator which was computationally tedious to obtain and gave almost no efficiency improvement. Table 1 summarizes the results based on 1000 simulation runs. It demonstrates that only the proposed estimators of the conditional and marginal controlled direct effect log odds ratio are guaranteed to be unbiased. As predicted by the theory, the logistic regression coefficient of X , adjusting for both M and L , is unbiased when L is not affected by X ; i.e. $\alpha = 0$. The one adjusting for M only is unbiased when L is not associated with Y ; i.e. $\eta_l = \eta_{xl} = 0$. Here, bias was calculated as the average of the obtained estimates minus the population value which equalled 0.25 for all estimators, except for the marginal log odds ratio where the population value was calculated analytically as in (6), yielding values varying between 0.20 and 0.25 over the different experiments. The third line in Table 1 corresponds to a setting where all estimators are unbiased. Interestingly, it demonstrates that unnecessary adjustment for L harms the efficiency of the standard logistic regression-based direct effect estimator, but not of the proposed estimators. Similar results were found for other parameter settings; not shown. The last three simulation experiments evaluated the impact of ignoring interactions between X and L and reveal a relatively mild bias when using the estimate of β in the misspecified model as an estimate of the direct effect at $L = E(L|X)$.

In a second simulation experiment, we evaluated the performance of the proposed estimators for case-control studies. We generated a large sample of measurements (X, L, M, Y) as before, but with X a Bernoulli variate with mean 0.5, $\eta_l = 1$ and $\eta_{xl} = 0$. We let β take the values 0 and 2, and chose λ to yield values of $E\{Y(0, 0)\}$ equalling 0.001, 0.01, 0.025, 0.05 and 0.1. We then randomly sampled 100 cases and 100 controls from these simulated data and next evaluated the previously considered standard logistic regression coefficients and the proposed estimator of the conditional controlled direct effect log odds ratio obtained by solving (13) using the computationally attractive choice (9). The results, based on 1000 simulation runs, are summarized in Table 2 and demonstrate again that only the proposed estimator is guaranteed to be approximately unbiased, with the bias slightly increasing with increasing population prevalence. The case where L is not affected by X , i.e. $\alpha = 0$, is of interest as it allows for comparing the precision between the estimators. It shows a minor efficiency advantage of the proposed estimator relative to standard logistic regression.

5. DISCUSSION

The results in this paper can straightforwardly be extended to the estimation of direct effect odds ratios, controlling the mediator at an arbitrary reference level m instead of zero, by considering models of the form

$$\frac{\text{odds}\{Y(x, m) = 1 \mid X = x, L(x)\}}{\text{odds}\{Y(0, m) = 1 \mid X = x, L(x)\}} = \exp\{\gamma(x; \beta_m^*)\}$$

for each m . Here, we allow for a different parameter β_m^* for each m to ensure that, for a given fixed law $f(L \mid X)$ and association model (5) evaluated at η^* , there exists a value β_m^* for β that makes (7) hold for all X and m . This does not induce a curse of dimensionality problem, even when M is absolutely continuous. In future work, we will examine how to test the null hypothesis of no exposure-mediator interaction, i.e., that $\beta_m^* = \beta_{m'}^*$ for all m, m' . In addition, we will investigate how to choose the association model (5) compatible with this null hypothesis

Table 2: Validity and efficiency of competing estimators of the controlled direct effect log odds ratio in case-control samples.

α	β	θ	LR-M				LR-ML				LD-C			
			Bias	ES	SS	Cov	Bias	ES	SS	Cov	Bias	ES	SS	Cov
1	2	0.1	0.45	5.3	5.1	90.7	-0.97	9.0	8.5	75.0	0.02	5.2	5.2	96.1
		1	0.45	5.2	4.9	89.0	-0.94	8.3	7.9	75.0	-0.03	5.2	5.0	93.6
		2.5	0.43	5.0	4.8	89.4	-0.96	8.1	7.7	73.5	-0.08	5.1	5.0	94.5
		5	0.39	4.9	4.7	89.2	-0.99	7.9	7.5	72.9	-0.15	5.1	4.9	92.8
		10	0.37	4.6	4.6	92.9	-1.03	7.6	7.4	69.3	-0.20	5.0	4.9	92.6
	0	0.1	4.41	4.5	4.4	86.7	-1.05	9.2	8.3	75.2	0.01	4.7	4.6	95.2
		1	0.40	4.4	4.2	85.1	-1.06	8.4	7.9	73.1	0.02	4.4	4.3	95.9
		2.5	0.42	4.2	4.1	84.7	-1.06	8.0	7.7	72.4	0.04	4.3	4.2	96.0
		5	0.42	4.1	4.0	83.0	-1.03	8.1	7.6	72.6	0.07	4.2	4.1	95.3
		10	0.39	4.1	3.9	83.2	-1.05	7.8	7.4	71.2	0.06	4.1	4.0	94.5
0	2	0.1	1.28	6.2	6.0	41.7	0.07	8.8	8.1	94.5	0.07	8.5	7.9	95.5
		1	1.27	6.1	5.8	39.5	0.08	8.1	7.8	95.3	0.04	7.9	7.6	95.6
		2.5	1.25	6.0	5.8	41.2	0.08	8.0	7.6	94.0	0.03	7.8	7.5	94.8
		5	1.22	6.0	5.7	43.8	0.08	7.9	7.5	94.3	0.02	7.8	7.4	93.7
		10	1.18	6.1	5.7	46.4	0.04	7.9	7.3	93.3	-0.03	7.8	7.3	93.5
	0	0.1	1.24	5.8	5.7	40.5	0.01	9.1	8.4	94.3	0.03	8.5	8.0	94.6
		1	1.23	5.3	5.3	34.5	0.01	8.3	7.9	94.1	0.07	7.8	7.4	95.0
		2.5	1.23	5.3	5.2	32.9	-0.01	8.0	7.7	95.6	0.08	7.4	7.1	95.4
		5	1.21	5.2	5.0	30.6	-0.03	8.0	7.5	93.8	0.09	7.3	6.8	94.3
		10	1.22	5.2	4.9	28.2	0.01	8.0	7.3	94.4	0.17	7.2	6.5	93.1

Empirical bias (multiplied by 10), empirical standard error (ES) (multiplied by 10), average of the sandwich standard error estimates (SS) (multiplied by 10) and coverage (in %) of 95% confidence intervals for the standard logistic regression estimates, adjusting for M (LR-M) and (M, L) (LR-ML), and of the proposed estimator of the conditional controlled direct effect log odds ratio (LD-C); $\theta = 100E\{Y(0, 0)\}$.

and will study the asymptotic distribution of the proposed estimators when a generalized additive association model is used.

ACKNOWLEDGEMENT

I am grateful to Rhian Daniel, Vanessa Didelez and Tyler VanderWeele, and to 3 referees and an Associate Editor for helpful discussions and comments.

APPENDIX 1

Proof of Proposition 1

That model $\mathcal{A}$ implies $\mathcal{B}$ is immediate. To prove the opposite, we must exhibit for any observed data law $f(Y, M, L, X)$ satisfying the restrictions of model $\mathcal{B}$ a full data law $f\{Y(X, 0), Y(0, 0), Y, M, L, X\}$ satisfying the restrictions of model $\mathcal{A}$. First, we set $f(Y, M, L, X)$ equal to an arbitrary joint law that satisfies (5). It follows from Gill & Robins (2001) that the consistency assumption and the no unmeasured confounders assumptions (2) impose no restrictions on the observed data law, so that we can always find a law $f\{Y(X, 0), Y(0, 0) \mid Y, M, L, X\}$ satisfying these restrictions. We can choose this law to satisfy

$E\{Y(0, 0) \mid L, X\} = \text{expit}[\text{logit}E\{Y(X, 0) \mid L, X\} - \gamma(X; \beta^*)]$ in order to make (4) hold. Indeed, this restriction does not conflict the consistency assumption by the fact that $Y(0, 0)$ is not observable, unless $X = 0$ in which case the above equality becomes redundant. It does not conflict assumption (3) because the above identity is a restriction on the law $f\{Y(0, 0) \mid L, X\}$ and therefore imposes no restrictions on the conditional dependence between $Y(0, 0)$ and M , given L and X . Finally, it does not conflict assumption (2) when model $\mathcal{A}$ satisfies (7) for all X .

APPENDIX 2

Linear structural direct effect models

Suppose that $E(Y \mid M, L, X) = \mu_{xl}(X, L; \eta^*) + \mu_m(X, L, M; \eta^*)$ and $E\{Y(X, 0) - Y(0, 0) \mid X, L\} = \gamma_x(X; \beta^*) + \gamma_l(X, L; \beta^*)$ with $\mu_{xl}(X, L; \eta^*) = \mu_x(X; \eta^*) + \mu_l(X, L; \eta^*)$, $\gamma_l(X, L; \beta)$ a known function of X and L , smooth in β and satisfying $\gamma_l(0, L; \beta) = 0$ and centered such that $E\{\gamma_l(X, L; \beta^*) \mid X\} = 0$, where $\gamma_x(X; \beta)$ and $\mu_m(X, L, m; \eta)$ are defined as before. Suppose that we fit the misspecified model $E\{Y(X, 0) - Y(0, 0) \mid X, L\} = \gamma_x(X; \beta)$ using the strategy of §2.3, but with the inverse logit link replaced with the identity link; i.e. by solving

$$0 = \sum_{i=1}^n d(M_i, L_i, X_i) \{Y_i - m(M_i, L_i, X_i; \eta)\} + [e(X_i) - E\{e(X_i)\}] \{\mu_{xl}(X_i, L_i; \eta) - \gamma_x(X_i; \beta)\}.$$

Upon setting $\eta = \eta^*$ and $\gamma_x(X; \beta) = \gamma_x(X; \beta^*)$, we obtain the following expected estimating equation $E\{[e(X_i) - E\{e(X_i)\}][E\{Y_i(0, 0) \mid X_i, L_i\} + \gamma_l(X_i, L_i; \beta^*)]\}$, which is zero because $E\{Y_i(0, 0) \mid X_i\} = E\{Y_i(0, 0)\}$ by (2) and because $E\{\gamma_l(X_i, L_i; \beta^*) \mid X_i\} = 0$. It follows that, if one were to ignore X - L interactions in the considered structural direct effect models with identity link, then the proposed estimators $\gamma_x(X; \hat{\beta})$ still converge to $E\{Y(X, 0) - Y(0, 0) \mid X\}$ in probability.

APPENDIX 3

Semi-parametric efficiency

We derive the efficient score for the parameters indexing model $\mathcal{B}$. To be sufficiently general, we relax (2) to be $Y(x, m) \perp\!\!\!\perp X \mid C$ for all x, m , where C is a collection of baseline covariates, and we redefine L such that $C \subseteq L$. Further, we redefine the association model (5) to be $\text{logit}E(Y \mid X, L, M) = \mu_x(X, C; \eta^*) + \mu_l(X, L; \eta^*) + \mu_m(X, L, M; \eta^*)$, where $\mu_x(X, C; \eta)$ is an arbitrary function of (X, C) , smooth in η , and the structural model (4) to be $\text{logit}E\{Y(X, 0) \mid X, L\} - \text{logit}E\{Y(0, 0) \mid X, L\} = \gamma(X, C; \beta^*)$, where $\gamma(X, C; \beta)$ is an arbitrary function of X and C , smooth in β and satisfying $\gamma(0, C; \beta) = 0$, and where β^* is an unknown, finite-dimensional parameter. For reasons explained in §2.3, one should ideally impose minimal restrictions on the term $\mu_x(X, C; \eta)$ in the association model.

The likelihood for a single observation equals $f(Y \mid M, L, X; \eta)f(M, L \mid X, C; \beta, \eta, \eta_{ml})f(X, C; \eta_x)$, where

$$\begin{aligned} & \int \text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} f(L \mid X, C; \beta, \eta, \eta_{ml}) dL \\ &= \int \text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} f(L \mid X, C; \beta, \eta, \eta_{ml}) f(X \mid C; \eta_x) dL dX, \end{aligned}$$

with $\mu_{xl}(X, L; \eta) = \mu_x(X, C; \eta) + \mu_l(X, L; \eta)$ and η_x and η_{ml} infinite-dimensional nuisance parameters. The tangent space (Newey, 1990) for η_x is given by $\Lambda_x = \{S(X, C) : E\{S(X, C)\} = 0\}$ and the tangent space for η_{ml} is given by

$$\begin{aligned} \Lambda_{ml} &= \{S(M, L \mid X, C) : E\{S(M, L \mid X, C) \mid X, C\} = 0, \\ &E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} S(M, L \mid X, C) \mid X, C] \\ &= E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} S(M, L \mid X, C) \mid C]\}. \end{aligned}$$

It is easy to verify that the projection of an arbitrary function $H \equiv h(Y, M, L, X)$ onto the orthocomplement of the nuisance tangent space Λ_{ml} is

$$\begin{aligned} H - E(H | M, L, X) + E(H | X, C) \\ + (\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} - E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} | C]) \\ \times \sigma^{-2}(X, C) \{\text{Cov}[H, \text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} | X, C] \\ - E(\sigma^{-2}(X, C) \text{Cov}[H, \text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} | X, C] | C) E\{\sigma^{-2}(X, C) | C\}^{-1}\}, \end{aligned}$$

where $\sigma^2(X, C) \equiv \text{var}[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} | X, C]$. The scores for η under parametric submodels can be written as

$$m_\eta(M, L, X; \eta) [Y - \text{expit}\{m(M, L, X; \eta)\}] + S_\eta(M, L | X, C) \quad (\text{A1})$$

where $m(M, L, X; \eta) = \mu_{xl}(X, L; \eta) + \mu_m(M, L, X; \eta)$ and where $S_\eta(M, L | X, C)$ satisfies $E\{S_\eta(M, L | X, C) | X, C\} = 0$ and

$$\begin{aligned} E[\text{expit}'\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} \mu_{xl, \eta}(X, L; \eta) | X, C] \\ + E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} S_\eta(M, L | X, C) | X, C] \\ = E[\text{expit}'\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} \mu_{xl, \eta}(X, L; \eta) | C] \\ + E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} S_\eta(M, L | X, C) | C]. \end{aligned} \quad (\text{A2})$$

The first term in (A1) is orthogonal to the nuisance tangent spaces Λ_x and Λ_{ml} . The projection of the second term onto the orthocomplement of the nuisance tangent space Λ_{ml} is given by

$$\begin{aligned} (\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} - E[\text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} | C]) \\ \times \sigma^{-2}(X, C) \{E[\text{expit}'\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} \mu_{xl, \eta}(X, L; \eta) | X, C] - E^{-1}\{\sigma^{-2}(X, C) | C\} \\ \times E(\sigma^{-2}(X, C) E[\text{expit}'\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\} \mu_{xl, \eta}(X, L; \eta) | X, C] | C)\} \end{aligned}$$

which is also orthogonal to Λ_x . The efficient score for η is thus given by the last q rows of the efficient estimating function given in §2.3. Likewise, the scores for β under parametric submodels can be written as $S_\beta(M, L | X, C)$ where $S_\beta(M, L | X, C)$ satisfies $E\{S_\beta(M, L | X, C) | X, C\} = 0$ and (A2) with $S_\eta(M, L | X, C)$ replaced by $S_\beta(M, L | X, C)$ and $\mu_{xl, \eta}(X, L; \eta)$ by $\gamma_\beta(X, C; \beta)$. Its projection onto the orthocomplement of the nuisance tangent space Λ_l is as given by the first p rows of the efficient estimating function given in §2.3, which is also orthogonal to Λ_x , and is thus the efficient score for β .

A locally robust estimator can be obtained by estimating β^* as the solution to an estimating equation based on the following estimating functions $[\gamma_\beta(X, C; \beta) - E\{\gamma_\beta(X, C; \beta) | C\}] \text{expit}\{\mu_{xl}(X, L; \eta) - \gamma(X, C; \beta)\}$, and η^* as the solution to an estimating equation based on a set of estimating functions which includes $[\gamma_\beta(X, C; \beta) - E\{\gamma_\beta(X, C; \beta) | C\}] [Y - \text{expit}\{m(X, L, M; \eta)\}]$. Indeed, let $\tilde{\eta}_0$ be the probability limit of $\hat{\eta}$ in the absence of a total effect; i.e., $\beta^* = 0$ and $\mu_m(M, L, X; \eta^*) = 0$. From the fact that the second estimating equation above is set to have mean zero, we obtain that $E([\gamma_\beta(X, C; \beta) - E\{\gamma_\beta(X, C; \beta) | C\}] \text{expit}\{\mu_{xl}(X, L; \tilde{\eta}_0)\}) = E\{[\gamma_\beta(X, C; \beta) - E\{\gamma_\beta(X, C; \beta) | C\}] Y\}$ in the absence of a total effect. It follows that the first estimating function above has mean zero at $\beta^* = 0$ and $\mu_m(M, L, X; \eta^*) = 0$ because its mean equals the conditional covariance between $\gamma_\beta(X, C; 0)$ and Y , given C , which is zero in the absence of a causal effect. It is easily verified that the choice (10) yields estimating functions that satisfy the aforementioned restrictions and thus yields a locally robust estimator.

REFERENCES

- BARON, R. & KENNY, D. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**, 1173–1182.
 COLE, S. & HERNÁN, M. (2002). Fallibility in estimating direct effects. *Int. J. Epidemiol.* **31**, 163–165.

- EMSLEY, R., DUNN, G. & WHITE, I. R. (2009). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat. Methods Med. Res.*, doi:10.1177/0962280209105014.
- GALLOP, R., SMALL, D., LIN, J., ELLIOTT, M., JOFFE, M. & TEN HAVE, T. (2009). Mediation analysis with principal stratification. *Statist. Med.* **28**, 1108–1130.
- GILL, R. & ROBINS, J. (2001). Causal inference for complex longitudinal data: The continuous case. *Ann. Statist.* **29**, 1785–1811.
- GOETGELUK, S., VANSTEELENDT, S. & GOETGHEBEUR, E. (2008). Estimation of controlled direct effects. *J. Royal Statist. Soc., Series B* **70**, 1049–1066.
- HERNÁN, M. & ROBINS, J. (2006). Instruments for causal inference an epidemiologist's dream? *Epidemiology* **17**, 360–372.
- IMAI, K., KEELE, L. & TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, in print.
- JOFFE, M. & GREENE, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65**, 530–538.
- JOFFE, M., SMALL, D. & HSU, C.-Y. (2007). Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statist. Science* **22**, 74–97.
- MACKINNON, D. (2008). *An Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum Associates.
- MIETTINEN, O. (1972). Components of the crude risk ratio. *Am. J. Epidem.* **96**, 168–172.
- NEWBY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Economet.* **5**, 99–135.
- PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- PEARL, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco: Morgan Kaufmann.
- PETERSEN, M., SINISI, S. & VAN DER LAAN, M. (2006). Estimation of direct causal effects. *Epidemiology* **17**, 276–284.
- PRENTICE, R. & PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403–411.
- ROBINS, J. (1999a). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, M. Halloran & D. Berry, eds. New-York: Springer-Verlag, pp. 95–134.
- ROBINS, J. (1999b). Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. In *Computation, Causation, and Discovery*, C. Glymour & G. Cooper, eds. Menlo Park, CA, Cambridge: AAAI Press/The MIT Press, pp. 349–405.
- ROBINS, J. & GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–155.
- ROBINS, J., HERNÁN, M. & BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.
- ROBINS, J. & ROTNITZKY, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* **91**, 763–783.
- ROSENBLUM, M., JEWELL, N. P., VAN DER LAAN, M., SHIBOSKI, S., VAN DER STRATEN, A. & PADIAN, N. (2009). Analysing direct effects in randomized trials with secondary interventions: an application to human immunodeficiency virus prevention trials. *Statist. Soc.* **172**, 443–465.
- RUBIN, D. (2004). Direct and indirect causal effects via potential outcomes. *Scand. J. Statist.* **31**, 161–170.
- TEN HAVE, T., JOFFE, M., LYNCH, K., BROWN, G., MAISTO, S. & BECK, A. (2007). Causal mediation analyses with rank preserving models. *Biometrics* **63**, 926–934.
- VAN DER LAAN, M. & PETERSEN, M. (2008). Direct effect models. *Int. J. Biostat.* **4**, Article 23.
- VANDERWEELE, T. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* **20**, 18–26.
- VANDERWEELE, T. & VANSTEELENDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2**, 457–468.
- VANDERWEELE, T. & VANSTEELENDT, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *Am. J. Epidem.* **171**, in print.
- VANSTEELENDT, S. (2009). Estimating direct effects in cohort and case-control studies. *Epidemiology* **20**, 851–860.
- VANSTEELENDT, S., GOETGELUK, S., LUTZ, S., WALDMAN, I., LYON, H., SCHADT, E., WEISS, S. & LANGE, C. (2009). On the adjustment for covariates in genetic association analysis: A novel, simple principle to infer causality. *Genet. Epidemiol.* **33**, 394–405.
- VANSTEELENDT, S. & GOETGHEBEUR, E. (2003). Causal inference with generalized structural mean models. *J. Royal Statist. Soc., Series B* **65**, 817–835.
- VANSTEELENDT, S. & GOETGHEBEUR, E. (2004). Using potential outcomes as predictors of treatment activity via strong structural mean models. *Statist. Sinica* **14**, 907–925.