Statistica Sinica **14**(2004), 907-925

USING POTENTIAL OUTCOMES AS PREDICTORS OF TREATMENT ACTIVITY VIA STRONG STRUCTURAL MEAN MODELS

S. Vansteelandt and E. Goetghebeur

Ghent University

Abstract: Structural mean models were developed to estimate average treatment effects in function of received exposures and baseline covariates. Recent extensions allow treatment effects to vary additionally with subjects' potential response to a treatment-free regime. This makes it possible to investigate in clinical trials, for instance, how well drug action is predicted by patients' natural health status in the absence of treatment. Accommodating this is challenging, however, because treatment activity and potential treatment-free response are (usually) unobservable for subjects on treatment.

In this paper, we model and estimate the effect of treatment-free outcomes on treatment activity in randomized controlled clinical trials with measured compliance. Our purpose is (a) to enhance modelling flexibility over existing approaches; and (b) to investigate to what extent the identification of such effects relies on untestable modelling assumptions. We develop new classes of estimators for this effect that make better use of the information in the data and achieve greater robustness to model misspecification. The new methods are evaluated by large sample approximation and a simulation study.

Key words and phrases: Causal inference, noncompliance, structural models.

1. Introduction

The quest for explanations of causal effects has continued through the centuries with ever more sophisticated tools. The randomized experiment, a milestone, was designed to study the impact of interventions when outcomes vary naturally between units. Its result is often summarized as an average intention-to-treat effect, even though most interventions have effects which themselves are likely to vary between units. Explaining such variation in effect has been an important but difficult goal.

Candidate prognostic factors for structural treatment effects come in at least three classes: I. the observed level/regime of experimental exposure, II. measurable baseline characteristics and III. potential intervention-free outcomes. The latter represent subjects' potential response to an intervention-free regime and are unmeasured for subjects who are exposed to the intervention. In a clinical

study, they quantify natural health status and form an adequate measure of comparison between patients. This is because potential intervention-free outcomes, like gender, are fixed characteristics of each subject, which are unaffected by earlier treatment.

Structural mean models (Robins (1994, 1997a, 1999) and Goetghebeur and Lapp (1997)) were developed and applied to explain causal effects in terms of Class I and Class II predictors. Class III predictors pose more challenging problems but remain of interest because, as we explain next, they may well account for much of the variation in treatment effect.

As a motivating example, we consider a randomized placebo-controlled blood pressure reduction trial described and analyzed in Goetghebeur and Lapp (1997). In a reanalysis, Fischer-Lapp and Goetghebeur (2001) observe that the lower quantiles of the diastolic blood pressure distribution at end-of-study are roughly the same in the treatment and placebo group, but that the median is lower in the treatment group. This suggests that treatment is effective but does not reduce blood pressures below the normal limits when those would be reached without treatment. Thus treatment is likely less effective for patients whose blood pressure normalizes in the absence of treatment. Also in general, interventions may be more effective among those who need it most, suggesting that the effect of exposure on outcome interacts with treatment-free response.

The difficulty with estimating the effects of treatment-free response on treatment activity stems from the lack of joint observations on observed treatment levels and potential intervention-free response. Efron and Feldman (1991) and Fischer-Lapp and Goetghebeur (2001) have tackled this problem in different ways in the context of randomized controlled trials with noncompliance. The former assumed observed compliance (i.e., exposure) levels to be exchangeable on a randomized treatment and control arm. The latter avoided this assumption but relied on a semiparametric compliance selection model in addition to a structural mean model. The purpose of this paper is (a) to develop new classes of estimators for these effects that make more efficient use of the information in the data and enhance robustness to model misspecification; and (b) to investigate to what extent the identification of such effects relies on untestable modelling assumptions.

The paper is organized as follows. In Section 2, we introduce the potential outcomes framework and structural mean models. In Section 3, we discuss causal models which enable treatment efficacy to depend upon treatment-free response. We then review the estimation strategy of Fischer-Lapp and Goetghebeur (2001) to estimate the parameters indexing those models. Next, we build on this approach to derive asymptotically efficient estimators for the average causal effect under more flexible and less restrictive models. In Section 4, we

show that the effect of treatment-free response on treatment activity cannot be identified without relying on untestable assumptions and we clarify the nature of those assumptions. Simulation results in Section 5 illustrate the finite sample performance of our estimators. We end with a discussion.

2. Potential Outcomes and Structural Mean Models

One popular definition of causal effects is cast in terms of counterfactual or potential outcomes (Rubin (1978)). Basically, one complements observed outcomes Y_i for subjects i $(i = 1, \ldots, n)$ with potential outcomes Y_{i0}. The latter indicate a reference response in the absence of exposure under similar experimental conditions. For instance, let $\mathbf{Z}_i$ be a vector measuring the observed level of exposure corresponding to observed outcome Y_i, such as treatment compliance as measured by the total number of pills taken by patient i in the blood pressure study (Fischer-Lapp and Goetghebeur (2001)). Then, the expected contrast

$$\mathrm{E}(Y_i - Y_{i0}|\mathbf{Z}_i), \tag{1}$$

defines the average causal effect of exposure level $\mathbf{Z}_i$ for subjects who were exposed to $\mathbf{Z}_i$. This is one possible causal parameter of interest.

Estimation of (1) is complicated by the lack of joint observations on Y_i and Y_{i0} (except in the absence of exposure). Identification of (1) is therefore not possible without making a number of untestable modelling assumptions that we list below. Most of these assumptions are justified by design if we focus on placebo-controlled double-blind randomized trials with an unexposed control arm (meaning that no treatment switches occur in that arm). In this setting, we let $\mathbf{Z}_i$ measure the level of *experimental* exposure and use as a reference outcome Y_{i0} the potential outcome of subject i following assignment to the control arm. For convenience, this will be denoted Y_{i0} as before and be called an intervention-free, treatment-free or exposure-free response. We make the following assumptions.

(A1) *Randomization assumption*: measurements are available for each subject on a randomization indicator R_i. This indicator is such that average exposure-free outcomes are independent of it, within strata of baseline covariates $\mathbf{x}_i$, i.e.,

$$\mathrm{E}(Y_{i0}|\mathbf{x}_i, R_i) = \mathrm{E}(Y_{i0}|\mathbf{x}_i).$$

We consider a two-arm study and let $R_i = 1$ indicate the experimental arm and $R_i = 0$ the control arm.

(A2) *Exclusion restriction* (Angrist, Imbens and Rubin (1996)): randomization R_i has no direct effect on the outcome (only an indirect effect via the exposure is possible). In double-blind randomized trials of an asymptomatic disease, one expects this to hold since patients and physicians are unaware of the assigned treatment (Robins (1994)).

(A3) *Uncontaminated control arm* (Cuzick, Edwards and Segnan (1997)): the control arm is unexposed. Letting $\mathbf{Z}_i = \mathbf{0}$ indicate absence of active exposure without loss of generality, this means that $\mathbf{Z}_i = \mathbf{0}$ when $R_i = 0$.

(A4) *Consistency assumption*: to link treatment-free outcomes to the observed data, we assume that $Y_i = Y_{i0}$ when $\mathbf{Z}_i = \mathbf{0}$. In that case we observe Y_{i0} in the control arm.

Finally, we assume that a model for the randomization probabilities $\text{pr}(R_i = 1|\mathbf{x}_i)$ is known, as is usually the case in randomized experiments.

For uncensored outcomes, one successful approach to modelling and estimating the causal effect (1), is based on the semiparametric structural mean model (SMM) of Robins (1994, 1997a). The simplest instance of such model is

$$\text{E}(Y_i - Y_{i0}|Z_i, R_i = 1) = \psi Z_i.$$

Here, ψ expresses the expected change in outcome when subjects who were exposed to $Z_i = 1$ would have their exposure set to zero. It explains randomization effects in terms of Class I predictors (i.e., received exposure). When also Class II predictors are important, one may add covariate-exposure interactions to the linear component of the model, as in

$$\text{E}(Y_i - Y_{i0}|Z_i, x_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i x_i,$$

for some baseline covariate x_i. Here, ψ_2 defines the change in the average effect of unit exposure per one-unit increase in x_i. Inference for this and more general types of structural mean models (sometimes involving time-dependent covariates), has been considered by Robins (1994, 1997a, 1999), Goetghebeur and Lapp (1997), Fischer-Lapp and Goetghebeur (1999), Robins and Rotnitzky (2003) and Vansteelandt and Goetghebeur (2003).

3. Strong Structural Mean Models

3.1. Review

Goetghebeur and Lapp (1998) and Fischer-Lapp and Goetghebeur (2001) argue that intervention-free response may be an important predictor of treatment activity. They incorporate interactions between univariate exposure measurements Z_i and intervention-free outcomes Y_{i0} via the strong structural mean model (SSMM)

$$\text{E}(Y_i - Y_{i0}|Y_{i0}, Z_i, \mathbf{x}_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i Y_{i0}. \tag{2}$$

In this model, $\psi_1 + \psi_2 Y_{i0}$ expresses how the observed outcome is expected to have changed from the latent treatment-free outcome for subjects with unit exposure,

treatment-free outcome Y_{i0} and covariate value $\mathbf{x}_i$. Within the same subgroup, ψ_1 measures the average causal effect of an observed unit dose when treatment-free outcome equals zero, and ψ_2 measures the difference in effect between subpopulations that differ one unit in treatment-free outcome.

To cope with the lack of joint observations on Y_i and Y_{i0}, Fischer-Lapp and Goetghebeur supplement the SSMM with a semiparametric compliance selection model. This model expresses how treatment-free outcomes have been 'selected' to receive certain exposure or compliance levels. In particular, they parameterize the residual association between treatment-free outcomes and observed exposure linearly, after adjusting for baseline covariates. That is,

$$\mathrm{E}(Y_{i0}|Z_i, \mathbf{x}_i, R_i = 1) - \mathrm{E}(Y_{i0}|\mathbf{x}_i) = \beta\{Z_i - \mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)\}, \tag{3}$$

for some unknown scalar β. Setting $\beta = 0$ is tantamount to assuming that $\mathbf{x}_i$ contains all risk factors for the outcome which also predict exposure (formally, this means that Y_{i0} and Z_i are conditionally mean independent given $\mathbf{x}_i$ and $R_i = 1$). In the blood pressure study, where Y measures blood pressure and Z_i the total dose taken, positive values of β indicate that for patients with the same baseline covariates, higher exposures are more likely seen among patients who would have high blood pressure in the absence of treatment.

Model (3) enables us to restate the righthand side of (2) in terms of observables:

$$\mathrm{E}(Y_i - Y_{i0}|Z_i, \mathbf{x}_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i \mathrm{E}(Y_{i0}|\mathbf{x}_i) + \psi_3 Z_i\{Z_i - \mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)\}, \tag{4}$$

where $\psi_3 = \psi_2\beta$. To estimate the mean components $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)'$, pseudo exposure-free outcomes are now defined for each subject:

$$H_i(\boldsymbol{\psi}) = Y_i - R_i\left[\psi_1 Z_i + \psi_2 Z_i \mathrm{E}(Y_{i0}|\mathbf{x}_i) + \psi_3 Z_i\{Z_i - \mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)\}\right]. \tag{5}$$

These coincide with the observed exposure-free outcomes in the control arm (by (A4)). In the experimental arm, they represent conditionally unbiased predictors (given $\mathbf{x}_i$) for the exposure-free outcomes when $\boldsymbol{\psi}$ equals the true value $\boldsymbol{\psi}_0$ (by (4)). This is because outcomes $H_i(\boldsymbol{\psi})$ have equal conditional means in both randomized arms, given $\mathbf{x}_i$ and $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, by (A1). Contrasting them between arms therefore leads to an unbiased estimating equation for $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, i.e.,

$$\sum_{i=1}^{n}\{\mathbf{d}_1(\mathbf{x}_i)H_i(\boldsymbol{\psi}) + \mathbf{d}_2(\mathbf{x}_i)\}\{R_i - \mathrm{pr}(R_i = 1|\mathbf{x}_i)\} = \mathbf{0} \tag{6}$$

with $\mathbf{d}_1(\mathbf{x}_i), \mathbf{d}_2(\mathbf{x}_i)$ arbitrary vectors of dimension three.

The above approach has made it possible to estimate interactions between exposure and treatment-free response in causal models, but relies on some assumptions which can be avoided.

1. The estimating functions defined in (6) involve unknown nuisance parameters $\mathrm{E}(Y_{i0}|\mathbf{x}_i)$ and $\mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)$. When these are replaced by consistent estimates, Slutsky's theorem shows that the solution $\hat{\boldsymbol{\psi}}$ to (6) is a consistent estimator of $\boldsymbol{\psi}$. When $\mathbf{x}_i$ is high-dimensional with many continuous components, it is however not possible to obtain globally consistent estimators with good moderate sample size performance for these expectations, due to the curse of dimensionality. Hence, smooth working models must be chosen for $\mathrm{E}(Y_{i0}|\mathbf{x}_i)$ and $\mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)$ which can then be fitted in the usual way in the control arm and the experimental arm, respectively (e.g., via least squares regression). There are however many instances where such models can be hard to specify, especially when the received dose is highly skewed, involves excess zeroes, or when the exposure measurement is itself high-dimensional. In the next sections, we avoid smooth models for $\mathrm{E}(Y_{i0}|\mathbf{x}_i)$ and $\mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)$ in the strong structural mean frame.
2. Combining (3) and (4) yields a model for expected outcomes in the experimental arm, as a function of observed exposure and covariates

$$\begin{aligned}\mathrm{E}(Y_i|Z_i, \mathbf{x}_i, R_i = 1) = \mathrm{E}(Y_{i0}|\mathbf{x}_i)(1 + \psi_2 Z_i) + \psi_1 Z_i \\ +(\beta + \psi_3 Z_i)\{Z_i - \mathrm{E}(Z_i|\mathbf{x}_i, R_i = 1)\}. \qquad (7)\end{aligned}$$

Assuming that the SSMM (2) holds, (7) allows one in principle to verify (3) (see also the next section). Because (6) ignores the additional information about $\boldsymbol{\psi}$ that is available through this conditional mean restriction, it entails an inefficient subset of all unbiased estimating equations for (ψ_1, ψ_2). In Section 4, we show that, based on this subset of estimating functions, it may be difficult or even impossible to locally identify $\boldsymbol{\psi}$ even when there is information for (ψ_1, ψ_2). In the next two sections, we derive the set of *all* unbiased estimating functions for (ψ_1, ψ_2) and a representation for the efficient estimating function under a less restrictive model.

3.2. Enhancing flexibility

We build on the existing development to consider inference for the more flexible model

$$\mathrm{E}(Y_i - Y_{i0}|Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 + Y_{i0}\boldsymbol{W}'_{i2}\boldsymbol{\psi}_2. \qquad (8)$$

Here $\boldsymbol{W}_{i1} = \boldsymbol{W}_1(\mathbf{Z}_i, \mathbf{x}_i) \in \mathbb{R}^{p_1}$ and $\boldsymbol{W}_{i2} = \boldsymbol{W}_2(\mathbf{Z}_i, \mathbf{x}_i) \in \mathbb{R}^{p_2}$ are design vectors which include all measured variables, $\mathbf{Z}_i$ and $\mathbf{x}_i$, that predict the causal effect of exposure; $\boldsymbol{\psi}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\psi}_2 \in \mathbb{R}^{p_2}$ are unknown structural parameters and we redefine $\boldsymbol{\psi} = (\boldsymbol{\psi}'_1, \boldsymbol{\psi}'_2)'$. By (A4), both design vectors $\boldsymbol{W}_{i1}$ and $\boldsymbol{W}_{i2}$ must be $\mathbf{0}$

(w.p.1) when $\mathbf{Z}_i = \mathbf{0}$. Model (8) becomes the simpler model (2) upon setting $\boldsymbol{W}_{i1} = \boldsymbol{W}_{i2} = Z_i$.

While inference for $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ indexing (8) is possible via straightforward extension of the methods of Section 3.1, we follow a different route for the aforementioned reasons. First, we replace (3) and the modelling assumptions for $\mathrm{E}(Y_{i0}|\mathbf{x}_i)$ and $\mathrm{E}(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)$ by a single restriction:

$$\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \boldsymbol{W}'_{i3}\boldsymbol{\delta}, \tag{9}$$

where $\boldsymbol{W}_{i3} = \boldsymbol{W}_3(\mathbf{Z}_i, \mathbf{x}_i) \in \mathbb{R}^{p_3}$ and $\boldsymbol{\delta} \in \mathbb{R}^{p_3}$ is an unknown structural parameter. Like (3), (9) expresses how treatment-free outcomes have been selected to receive certain exposure levels. Choices of $\boldsymbol{W}_{i3}$ that depend solely on $\mathbf{x}_i$ formalize the no-unmeasured-confounders assumption (Robins (1997a)) that $\mathbf{x}_i$ contains all risk factors for the outcome that also predict exposure. More general choices express the fact that the residual association between exposure and outcome is confounded by unmeasured factors, even after adjustment for measured baseline covariates.

Model (9) easily enables one to relate treatment-free outcomes nonlinearly to exposure. Furthermore, it imposes fewer restrictions on the observed data law than the combination of (3) and models for $\mathrm{E}(Y_{i0}|\mathbf{x}_i)$ and $\mathrm{E}(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)$. This is seen by noting that the latter combination implies a model for $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$, namely $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \mathrm{E}(Y_{i0}|\mathbf{x}_i) + \boldsymbol{\beta}'\{\mathbf{Z}_i - \mathrm{E}(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)\}$, while the reverse is not true.

Model (9) may appear difficult to specify, but this is only seemingly so. In fact, when combined with (8), (9) yields a standard model for expected outcomes given observed exposure and baseline covariates in the experimental arm. The parameters of the latter model (including $\boldsymbol{\delta}$) are estimable in the usual way. Given (8), specifying (9) is thus essentially equivalent to specifying a model for expected outcomes given observed exposure and baseline covariates in the experimental arm.

Assuming that (8) and (9) hold with $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ and $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, we now develop the following formal estimation strategy. Together with (9), (8) yields pseudo exposure-free outcomes for subjects on treatment:

$$H_i(\boldsymbol{\psi}, \boldsymbol{\delta}) = Y_i - R_i\left(\boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2\boldsymbol{W}'_{i3}\boldsymbol{\delta}\right). \tag{10}$$

For given $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, a consistent point estimator for $\boldsymbol{\psi}$ can now be obtained − as before, by solving (6) − as the value which yields no mean difference in these outcomes between randomized arms (Robins (1994)). Similarly, combination of (8) and (9) yields a traditional model for expected outcomes given observed exposure:

$$\mathrm{E}(Y_i|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1; \boldsymbol{\psi}, \boldsymbol{\delta}) = \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 + (1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)\boldsymbol{W}'_{i3}\boldsymbol{\delta}. \tag{11}$$

For given $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, a consistent estimator for $\boldsymbol{\delta}$ in this model can be obtained by (weighted) least squares regression of outcomes on exposure and covariates in the experimental arm. The fact that such regression involves the unknown $\boldsymbol{\psi}_0$ suggests the following iteration scheme for estimating $\boldsymbol{\psi}_0$.

Step 1. Choose a starting value $\hat{\boldsymbol{\psi}}^{(0)}$ for $\boldsymbol{\psi}_0$, e.g., as obtained from Section 3.1.

Step 2. Given current estimates $\hat{\boldsymbol{\psi}}^{(k)}$, $k = 0, 1, 2, \ldots$, for $\boldsymbol{\psi}_0$, estimate $\boldsymbol{\delta}_0$ by fitting (11) to the observed outcomes in the active treatment arm, with $\hat{\boldsymbol{\psi}}^{(k)}$ in place of $\boldsymbol{\psi}$. This yields estimators $\hat{\boldsymbol{\delta}}^{(k+1)}$ for the unknown $\boldsymbol{\delta}_0$.

Step 3. Given current estimates $\hat{\boldsymbol{\delta}}^{(k)}$, $k = 1, 2, \ldots$, for $\boldsymbol{\delta}_0$, calculate pseudo exposure-free outcomes $H_i(\boldsymbol{\psi}, \hat{\boldsymbol{\delta}}^{(k)})$ by evaluating (10). Next, solve (6) for $\boldsymbol{\psi}$ with $\mathbf{d}_1(.), \mathbf{d}_2(.)$ arbitrary $(p_1 + p_2)$-dimensional functions of $\mathbf{x}_i$. This yields estimates $\hat{\boldsymbol{\psi}}^{(k+1)}$ for $\boldsymbol{\psi}_0$.

Step 4. Iterate steps 2 and 3 until $\|\hat{\boldsymbol{\psi}}^{(k+1)} - \hat{\boldsymbol{\psi}}^{(k)}\|$ is sufficiently small.

In the Appendix, we show that this algorithm yields regular, asymptotically linear (RAL) estimators for $\boldsymbol{\psi}_0$ (and hence consistent, asymptotically normal estimators for $\boldsymbol{\psi}_0$) under mild regularity conditions. More generally, we show that all regular asymptotically linear (RAL) estimators of $(\boldsymbol{\psi}_0, \boldsymbol{\delta}_0)$ are asymptotically equivalent to the solution of an estimating equation of the form

$$\sum_{i=1}^{n} \{R_i - \mathrm{pr}(R_i = 1|\mathbf{x}_i)\}\{\mathbf{d}_1(\mathbf{x}_i) H_i(\boldsymbol{\psi}, \boldsymbol{\delta}) + \mathbf{d}_2(\mathbf{x}_i)\}$$
$$+\mathbf{d}_3(\mathbf{Z}_i, \mathbf{x}_i) R_i \{Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 - (1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)\boldsymbol{W}'_{i3}\boldsymbol{\delta}\} = \mathbf{0} \qquad (12)$$

for some $(p_1+p_2+p_3)$-dimensional function $\mathbf{d}_1(.), \mathbf{d}_2(.)$ of $\mathbf{x}_i$ and $\mathbf{d}_3(.)$ of $(\mathbf{Z}_i, \mathbf{x}_i)$. We further derive an optimal choice $\mathbf{d}_{1opt}(.), \mathbf{d}_{2opt}(.), \mathbf{d}_{3opt}(.)$ for which the solution to (12) attains the semiparametric variance bound for $(\boldsymbol{\psi}_0, \boldsymbol{\delta}_0)$ in our model.

3.3. Enhancing robustness

So far, we have seen that flexible estimation strategies for SSMM's can be derived which avoid a combination of models for $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ and $\mathrm{E}(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)$. Instead, a single 'nuisance' model is used for the residual association between potential exposure-free outcomes and exposure after adjusting for baseline covariates, i.e., $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$. Misspecification of this model does not affect tests of the important null hypothesis of no average causal effect are robust to it. Furthermore, goodness-of-fit can be verified from (11) when (8) holds. Bad news enters, however, because minor misspecifications could yield inconsistent estimates in other points of the structural parameter space. This is

of special concern when there is little power to detect anomalies in the nuisance model.

To overcome this problem, we now consider inference for $\boldsymbol{\psi}$ under the sole SSMM assumption (8) (and the causal assumptions of Section 2). From (8), it follows that

$$\mathrm{E}(Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \mathrm{E}\{(1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1\},$$

and hence $\mathrm{E}((Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1)/(1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$. It follows that

$$H_i(\boldsymbol{\psi}_0) = \frac{Y_i - R_i\boldsymbol{W}'_{i1}\boldsymbol{\psi}_1}{1 + R_i\boldsymbol{W}'_{i2}\boldsymbol{\psi}_2}, \tag{13}$$

represents a pseudo exposure-free outcome. In the control arm, $H_i(\boldsymbol{\psi}_0)$ coincides with the observed exposure-free outcome (by (A4)). In the experimental arm, it represents a conditionally unbiased predictor (given $\mathbf{x}_i$) for the exposure-free outcome when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, by the fact that

$$\mathrm{E}\{H_i(\boldsymbol{\psi}_0)|\mathbf{x}_i, R_i = 1\} = \mathrm{E}\{H_i(\boldsymbol{\psi}_0)|\mathbf{x}_i, R_i = 0\}.$$

In the Appendix, we show that RAL estimators for $\boldsymbol{\psi}_0$ can be obtained (under regularity conditions) by solving (6) with pseudo-exposure free outcomes (13), provided $(Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1)/(1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)$ has finite variance in the experimental arm when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. We further show that (6) constitutes essentially all unbiased estimating equations for $\boldsymbol{\psi}_0$ under the observed data model defined by the sole restriction (8) and the assumptions of Section 2. That is, all RAL estimators for $\boldsymbol{\psi}_0$ under this model are asymptotically equivalent to the solution of (6) for some choice of $\mathbf{d}_1(.), \mathbf{d}_2(.)$. We further derive an optimal choice $\mathbf{d}_{1,opt}(.), \mathbf{d}_{2,opt}(.)$ for which the solution to (6) attains the semiparametric efficiency bound for our model.

The condition that $(Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1)/(1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)$ has finite variance in the experimental arm (when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$) is not easy to verify because it involves the unknown $\boldsymbol{\psi}_0$. Failure of this constraint is however detectable because it leads to severely inflated standard errors for $\hat{\boldsymbol{\psi}}$. In practice, it is sufficient for this condition to hold that $\boldsymbol{W}'_{i2}\boldsymbol{\psi}_2$ falls outside the interval $[-1 - \sigma, -1 + \sigma]$ w.p.1 for some $\sigma > 0$.

If the variance of $(Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1)/(1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)$ is high in the experimental arm, the estimating functions can become very unstable. Assuming that $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ is continuous in the points $\{(\mathbf{Z}_i, \mathbf{x}_i) : \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2 = 1\}$ when $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, greater stability may be obtained by replacing $H_i(\boldsymbol{\psi}_0)$ by a prediction

$$\mathrm{E}\left[\frac{Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1}{1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1\right] \tag{14}$$

for all observations i for which $\boldsymbol{W}'_{i2}\boldsymbol{\psi}_2 \in [-1-\sigma, -1+\sigma]$ (for some chosen $\sigma > 0$). These predictions can be obtained in the usual way via an ordinary least squares estimator or some kernel estimator based on all remaining observations. As the rule for selecting observations depends on the unknown $\boldsymbol{\psi}_0$, (6) must be solved through an iterative process. Alternatively, one may choose to delete observations i for which $\boldsymbol{W}'_{i2}\hat{\boldsymbol{\psi}}_2 \in [-1-\sigma, -1+\sigma]$ for some chosen $\sigma > 0$ and preliminary estimator $\hat{\boldsymbol{\psi}}_2$, and then update iteratively. The impact of deleting these observations or misspecifying a model for (14) will be negligible when the event $\boldsymbol{W}'_{i2}\boldsymbol{\psi}_2 \in [-1-\sigma, -1+\sigma]$ is rare under the truth $\boldsymbol{\psi} = \boldsymbol{\psi}_0$.

4. Identifiability

While modelling the impact of treatment-free outcomes on treatment activity was considered as a natural extension of the existing structural mean model methodology, doing this is unfortunately not possible without reliance on untestable assumptions. In this section, we investigate the nature of these assumptions.

Consider the following simple example in which x_i is a dichotomous covariate measured at baseline. Consider the semiparametric models $\mathcal{A}(1)$, $\mathcal{A}(2)$ and $\mathcal{A}(3)$ for the observed data, defined by the assumptions of Section 2 and an additional assumption given by one of the following models.

$$\mathcal{A}(1) : \mathrm{E}(Y_i - Y_{i0}|Z_i, Y_{i0}, x_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i Y_{i0}.$$
$$\mathcal{A}(2) : \mathrm{E}(Y_i - Y_{i0}|Z_i, Y_{i0}, x_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i x_i.$$
$$\mathcal{A}(3) : \mathrm{E}(Y_i - Y_{i0}|Z_i, Y_{i0}, x_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i Y_{i0} + \psi_2 Z_i x_i.$$

With a single dichotomous covariate x_i, there are only two identifying restrictions (6), hence at most two parameters can be identified in these models. In particular, all three models are nonparametric models for the observed data. It follows that, no matter how large the sample size, there is no empirical evidence favouring one model over the others. We conclude for this example that there is no information to disentangle the effects of covariates x_i and of treatment-free response Y_{i0} on treatment activity.

One possibility to cope with this identification problem, is to assume that the model holds within sufficiently rich covariate strata, i.e., that

$$\mathrm{E}(Y_i - Y_{i0}|Z_i, Y_{i0}, \mathbf{x}_i, R_i = 1) = \psi_1 Z_i + \psi_2 Z_i Y_{i0}. \tag{15}$$

In that case, one can solve (6) with richer $\mathbf{x}$-dependent indices $\mathbf{d}_1(\mathbf{x}), \mathbf{d}_2(\mathbf{x})$. The following theorem, proved in the Appendix, shows however that this strategy also relies on untestable assumptions.

Theorem 1. *Consider the semiparametric model $\mathcal{A}$ for the observed data, defined by the assumptions of Section 2 and the additional assumptions $E(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = h_1(\mathbf{Z}_i, \mathbf{x}_i)$ and $E(Y_i - Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, Y_{i0}, R_i = 1) = h_2(\mathbf{x}_i)Z_{i1} + q(\mathbf{Z}_i, \mathbf{x}_i)Y_{i0}$ with $h_1(.)$ and $h_2(.)$ unrestricted, $q(.)$ fixed and known satisfying $q(\mathbf{0}, \mathbf{x}_i) = 0, \forall \mathbf{x}_i$, and Z_{i1} the first scalar component of $\mathbf{Z}_i$. Then $\mathcal{A}$ defines a nonparametric identified model (Robins (1997b)) for the observed data provided that for each $\mathbf{x}_i$, $E[\mathbf{Z}_i/\{1+ q(\mathbf{Z}_i, \mathbf{x}_i)\} \mid \mathbf{x}_i, R_i = 1]$ differs from 0, and for each $(\mathbf{Z}_i, \mathbf{x}_i)$, $q(\mathbf{Z}_i, \mathbf{x}_i)$ differs from -1.*

Theorem 1 states that for given $q(.)$, the unknown parameters $h_1(.)$ and $h_2(.)$ can be identified. For unrestricted $h_1(.)$ and $h_2(.)$, it defines a nonparametric model for the observed data, meaning that it fits the observed data perfectly well. As such there can never by any data evidence favouring one value of $q(.)$ over another. In particular, it is not possible to identify nonparametrically whether intervention-free response predicts treatment activity (that is, whether or not the estimated effect is due to the independent effect of covariates alone). Because of this inherent identification problem, we argue that strong structural mean models should not be used as the single model in a data analysis, but should be compared with others in a sensitivity analysis. The results should be used in a hypothesis-generating manner, aimed to increase our understanding of the treatment mechanism.

Interestingly, our effort to make more efficient use of the information in the data, via the estimation methods of Sections 3.2 and 3.3, is further rewarded by a greater capacity to identify the unknown model parameters and thus less reliance on untestable modelling assumptions. To illustrate this, we consider the special case where x_i can take three different values and the model $\mathrm{E}(Y_i - Y_{i0}|Z_i, x_i, R_i = 1) = (\psi_1 + \psi_2 x_i + \psi_3 Y_{i0})Z_i$ holds. Fitting the incorrect model (2) with the methods of Section 3.1 would require the estimation of three unknown parameters, and thus leave no power to detect model misspecification. In contrast, the new methods generally allow identification of all three parameters ψ_1, ψ_2 and ψ_3 and thus enable us to detect that (2) is incorrectly specified. Likewise, consider the special case where x_i is dichotomous. Then the unknown parameters in (2) are inestimable with the methods of Section 3.1, but can generally be estimated with the new methods.

5. Simulation Results

A simulation study illustrates the finite sample performance of the proposed estimators. We generate 1,000 data sets of 1,000 independent samples as in Fisher-Lapp and Goetghebeur (2001). Standard uniformly distributed covariates $x_{i1} = U_i(0, 1)$ and standard normally distributed covariates $x_{i2} = N_i(0, 1)$ are

independently generated and subsequently used in all simulations. Conditional on them, exposure measurements $\tilde{Z}_i$ are generated as $\tilde{Z}_i = 0.5x_{i1} + 0.5U_i(0,1)$, and exposure-free outcomes obey the following working models.

1. $Y_{i0}|\tilde{Z}_i, x_{i1}, x_{i2} = 6 + 2x_{i1} - 2x_{i2} + 3\tilde{Z}_i + N_i(0,1)$.
2. $Y_{i0}|\tilde{Z}_i, x_{i1}, x_{i2} = 6 + 2x_{i1} - 2x_{i2} + 10\tilde{Z}_i^2 + N_i(0,1)$.
3. $Y_{i0}|\tilde{Z}_i, x_{i1}, x_{i2} = 6 + 2x_{i1} - 2x_{i2} - 5\tilde{Z}_i + 3\tilde{Z}_i x_{i1} + 3\tilde{Z}_i x_{i2} + 10\tilde{Z}_i^2 + N_i(0,1)$.

A randomization indicator R_i is generated independently of all other variables with success probability 0.5 and Z_i is taken to be $R_i\tilde{Z}_i$. Observed outcomes in the experimental arm ($R_i = 1$) are generated as

$$Y_i = Y_{i0} + \psi_1 Z_i + \psi_2 Z_i Y_{i0} + \epsilon_i, \tag{16}$$

with $\psi_1 = 5, \psi_2 = 1$, where ϵ_i $(i = 1, \ldots, n)$ are conditionally uncorrelated, mean zero random variables, given $(Z_i, Y_{i0}, \mathbf{x}_i, R_i = 1)$, with constant variance $\sigma_\epsilon^2 = 2$ when $\tilde{Z}_i \neq 0$ and $\sigma_\epsilon^2 = 0$ when $\tilde{Z}_i = 0$. In all simulations, estimates for the mean components are obtained via the estimation methods (A) of Section 3.1, (B) of Section 3.2 with $\mathrm{E}(Y_{i0}|\tilde{Z}_i, x_{i1}, x_{i2})$ linear in $\tilde{Z}_i, x_{i1}$ and x_{i2}, (C) of Section 3.3, and (D) of Section 3.2 with the correct working model for $\mathrm{E}(Y_{i0}|\tilde{Z}_i, x_{i1}, x_{i2})$. Optimal indices $\mathbf{d}_{1opt}(.), \mathbf{d}_{2opt}(.), \mathbf{d}_{3opt}(.)$ in the estimating equations were estimated by assuming linear regression models for all conditional expectations in the expressions for the efficient estimating function (see the Appendix) and by assuming that all conditional variances are independent of baseline covariates. For simplicity, we used the iterative algorithm of Section 3.2 in parts (B) and (D). Here, optimal indices were calculated for the separate estimating equations (i.e., efficient estimators were used for $\boldsymbol{\psi}$ assuming that $\boldsymbol{\delta}$ is known, and for $\boldsymbol{\delta}$ assuming that $\boldsymbol{\psi}$ is known). As such, we achieved reasonable but not full efficiency in parts (B) and (D). Table 1 summarizes the results.

Table 1. Simulation results: empirical bias and standard error based on 1,000 simulations of size 1,000.

Working model	Estimation method	Bias($\hat{\psi}_1$)	Bias($\hat{\psi}_2$)	SE($\hat{\psi}_1$)	SE($\hat{\psi}_2$)
1	A	-0.0243	2.30 E-3	1.014	0.109
	B	-0.0580	6.20 E-3	1.061	0.115
	C	-0.0590	6.36 E-3	1.090	0.118
2	A	-0.631	0.0605	1.623	0.147
	B	-0.489	0.0451	1.475	0.134
	C	-0.178	0.0193	1.483	0.136
	D	-0.182	0.0197	1.471	0.135
3	A	-1.454	0.166	1.653	0.182
	B	-0.963	0.110	1.494	0.168
	C	-0.268	0.0406	1.501	0.169
	D	-0.283	0.0421	1.514	0.171

We find that the three methods yield approximately unbiased estimates for the mean parameters when the conditional mean of Y_{i0} is linear in Z_i conditional on $\mathbf{x}_i$. The method of Section 3.1 has the smallest finite-sample bias for all parameters. Only the robust estimator of Section 3.3 and the estimator of Section 3.2 (with correct nuisance model) yield approximately unbiased estimators of the mean components when this assumption fails. All methods yield almost equally precise estimators for the mean components, suggesting that increased robustness to model misspecification need not necessarily come with information loss. For the mean components estimated via the robust estimator of Section 3.3, we have observed important efficiency improvements by implementing optimal indices in the estimating equations.

6. Discussion

Strong structural mean models (SSMM's) were introduced by Fischer-Lapp and Goetghebeur (2001) to model the dependence of treatment efficacy on potential treatment-free outcomes in a biologically meaningful way. In this paper, we have shown that this dependence is unfortunately not nonparametrically identified, meaning that identification relies on untestable assumptions. Despite this, we believe that the methods are useful if carefully applied. This is because treatment-free response is a useful reference indicator of natural health status (in the absence of treatment) and thus a possibly strong predictor of treatment activity. To avoid confounding the effect of covariates with the effect of treatment-free outcome, sufficient care must be exercised. In particular, we recommend always testing for the independent effects of covariates that are used for conditioning in (8).

In Section 4, we have shown that the parameters indexing SSMM's can be identified nonparametrically if one has available knowledge regarding certain non-identifiable functions $q(.)$. The choice $q(.) = 0$ encodes the investigator's belief that the average causal effect in the exposed is linear in the observed exposure and does not depend on treatment-free outcomes. Values of $q(.)$ different from 0 measure deviations from this extreme scenario. In general, however, the non-identifiable functions $q(.)$ are difficult to interpret. This makes it hard to perform a sensitivity analysis which evaluates the impact that different degrees of departure from $q(.) = 0$ have on inference about the average causal effect. It is an open question whether there exist possible reparameterizations of the model that could better enable such sensitivity analysis. Challenging extensions are also foreseen in accommodating this work to complex longitudinal data structures.

In summary, we have built on the work of Fischer-Lapp and Goetghebeur (2001) to derive new estimation methods for SSMM's that allow greater flexibility in model building and enhance robustness against misspecification of the

nuisance models. Simulation studies have shown that the new methods perform considerably better under such model misspecification. The new methods tend to be less stable in small samples, but their increased robustness does not appear to come at the expense of information loss in the moderate samples sizes encountered in practice.

Acknowledgement

We are grateful to an associate editor and a referee for insightful and constructive comments which greatly improved an earlier version of the manuscript. The first author acknowledges support as a Research Assistant and Postdoctoral Fellow of the Fund for Scientific Research - Flanders (Belgium) (F.W.O.),

Appendix

A.1. Semiparametric efficiency in Section 3.2

Consider Model $\mathcal{A}$ for the observed data, defined by (8), (9), (A1)−(A4) and $\text{pr}(R_i = 1|\mathbf{x}_i)$ known. Consider Model $\mathcal{A}^*$ for the observed data defined by the conditional mean models:

$$\text{E}\{U_{i1}(\boldsymbol{\psi}_0, \boldsymbol{\delta}_0)|\mathbf{x}_i\} = \text{E}(U_{i3}|\mathbf{x}_i) = 0 \tag{17}$$

$$\text{E}\{U_{i2}(\boldsymbol{\psi}_0, \boldsymbol{\delta}_0)|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1\} = 0, \tag{18}$$

where $U_{i1}(\boldsymbol{\psi}, \boldsymbol{\delta}) = \{Y_i - R_i(\boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2\boldsymbol{W}'_{i3}\boldsymbol{\delta})\}\{R_i - \text{pr}(R_i = 1|\mathbf{x}_i)\}, U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}) = Y_i - \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 - (1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)\boldsymbol{W}'_{i3}\boldsymbol{\delta}$ and $U_{i3} = R_i - \text{pr}(R_i = 1|\mathbf{x}_i)$. Then $\mathcal{A}$ and $\mathcal{A}^*$ define the same semiparametric model for the observed data. Indeed, it is easy to see that $\mathcal{A}^*$ is implied by the restrictions of $\mathcal{A}$. To show the reverse, we prove that for any observed data law satisfying the restrictions in $\mathcal{A}^*$, we can always exhibit a full data law satisfying $\mathcal{A}$ which marginalizes to this observed data law. Consider thus an arbitrary observed data law satisfying the restrictions for $\mathcal{A}^*$, defined by $f(\mathbf{x}_i) = f^*(\mathbf{x}_i)$, $\text{E}(R_i|\mathbf{x}_i) = \text{pr}^*(R_i = 1|\mathbf{x}_i)$, with $f(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1) = f^*(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1), \text{E}(Y_i|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \boldsymbol{W}'_{i1}\boldsymbol{\psi}_1 + (1 + \boldsymbol{W}'_{i2}\boldsymbol{\psi}_2)\boldsymbol{W}'_{i3}\boldsymbol{\delta}, f(\epsilon_{i1}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = f^*(\epsilon_{i1}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$, $\text{E}(Y_i|\mathbf{x}_i, R_i = 0) = \text{E}^*(\boldsymbol{W}'_{i3}\boldsymbol{\delta}|\mathbf{x}_i, R_i = 1)$ and $f(\epsilon_{i0}|\mathbf{x}_i, R_i = 0) = f^*(\epsilon_{i0}|\mathbf{x}_i, R_i = 0)$, for $\epsilon_{i1} = Y_i - \text{E}(Y_i|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$, $\epsilon_{i0} = Y_i - \text{E}(Y_i|\mathbf{x}_i, R_i = 0)$ and arbitrary well-defined densities $f^*(.)$ (and expectations $\text{E}^*(.)$). Then we construct a full data law that satisfies $\mathcal{A}$ and marginalizes to this observed data law. Choose $f(\mathbf{x}_i) = f^*(\mathbf{x}_i)$, $\text{E}(R_i|\mathbf{x}_i) = \text{pr}^*(R_i = 1|\mathbf{x}_i)$ and let $f(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1) = f^*(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)$. Further, let the conditional density of Y_{i0} given $(\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ be such that $\text{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = \boldsymbol{W}'_{i3}\boldsymbol{\delta}$ and $\int f(\epsilon_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)f^*(\mathbf{Z}_i|\mathbf{x}_i, R_i = 1)d\mathbf{Z}_i = f^*(\epsilon_{i0}|\mathbf{x}_i, R_i = 0)$. This is always possible upon choosing $Y_{i0} = \boldsymbol{W}'_{i3}\boldsymbol{\delta} + \nu_i$ with $f(\nu_i = x|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) = f(\nu_i = x|\mathbf{x}_i, R_i = 1) = f^*(\epsilon_{i0} = x|\mathbf{x}_i, R_i = 0)$. Finally, let the conditional density

of Y_i given $(Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ be such that $\mathrm{E}(Y_i|Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ is as given by (8) and $\int f(\epsilon_{i1}|Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1) f(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1) dY_{i0} = f^*(\epsilon_{i1}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$. Doing this is again possible using similar arguments as before.

We thus conclude that $\mathcal{A}$ and $\mathcal{A}^*$ define the same semiparametric model for the observed data. Further, the finite-dimensional parameter $(\boldsymbol{\psi}_0', \boldsymbol{\delta}_0')'$ indexing $\mathcal{A}$ and $\mathcal{A}^*$, which can be defined as the solution of (17)−(18), is the same functional of the observed data law under both models. It follows that all consistent asymptotically normal (CAN) estimators of $(\boldsymbol{\psi}_0', \boldsymbol{\delta}_0')'$ in Model $\mathcal{A}$ are CAN estimators of $(\boldsymbol{\psi}_0', \boldsymbol{\delta}_0')'$ in Model $\mathcal{A}^*$, and vice-versa. In particular, the semiparametric efficiency bound is the same for both models. Knowing this is important because it implies that semiparametric efficient estimating equations for $(\boldsymbol{\psi}_0', \boldsymbol{\delta}_0')'$ in $\mathcal{A}$ can be constructed as for $\mathcal{A}^*$.

Following Chamberlain (1992), the semiparametric efficiency bound for $\mathcal{A}^*$ is attained by solving an estimating equation of the form

$$\sum_i \mathbf{d}_1(\mathbf{x}_i) U_{i1}(\boldsymbol{\psi}, \boldsymbol{\delta}) + \mathbf{d}_2(\mathbf{Z}_i, \mathbf{x}_i) U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}) + \mathbf{d}_3(\mathbf{x}_i) U_{i3} = \mathbf{0}$$

for particular choices of $\mathbf{d}_1(\mathbf{x}_i), \mathbf{d}_2(\mathbf{Z}_i, \mathbf{x}_i)$ and $\mathbf{d}_3(\mathbf{x}_i)$. Furthermore, all RAL estimators of $(\boldsymbol{\psi}_0', \boldsymbol{\delta}_0')'$ are asymptotically equivalent to the solution of an estimating equation of this form, for some choice of $\mathbf{d}_1(\mathbf{x}_i), \mathbf{d}_2(\mathbf{Z}_i, \mathbf{x}_i), \mathbf{d}_3(\mathbf{x}_i)$. Using Chamberlain (1992), we find that the semiparametric variance bound for our model is attained by choosing:

$$\begin{aligned}
U_{i1}^* &:= U_{i1}(\boldsymbol{\psi}, \boldsymbol{\delta}) - \frac{\mathrm{E}\left\{U_{i1}(\boldsymbol{\psi}, \boldsymbol{\delta}) U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\right\}}{\mathrm{E}\left\{U_{i2}^2(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\right\}} U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}), \\
\mathbf{d}_1(\mathbf{x}_i) &= \frac{\mathrm{E}\{\nabla_\gamma U_{i1}^*(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{x}_i\}}{\mathrm{E}\{U_{i1}^{2*}(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{x}_i\}}, \\
\mathbf{d}_2(\mathbf{Z}_i, \mathbf{x}_i) &= \frac{\mathrm{E}\{\nabla_\gamma U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\}}{\mathrm{E}\{U_{i2}^2(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\}} \\
&\quad - \mathbf{d}_1(\mathbf{x}_i) \frac{\mathrm{E}\left\{U_{i1}(\boldsymbol{\psi}, \boldsymbol{\delta}) U_{i2}(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\right\}}{\mathrm{E}\left\{U_{i2}^2(\boldsymbol{\psi}, \boldsymbol{\delta}) | \mathbf{Z}_i, \mathbf{x}_i, R_i = 1\right\}}
\end{aligned}$$

and $\mathbf{d}_3(\mathbf{x}_i) = -\mathbf{d}_1(\mathbf{x}_i)\mathrm{E}(Y_i|R_i = 0, \mathbf{x}_i)$, where $\boldsymbol{\gamma} = (\boldsymbol{\psi}', \boldsymbol{\delta}')'$.

A.2. Semiparametric efficiency in Section 3.3

Suppose we consider the semiparametric model $\mathcal{A}_{sub}$ defined by (8), the causal assumptions of Section 2 with $\mathrm{pr}(R_i = 1|\mathbf{x}_i)$ known and arbitrary correctly specified parametric submodel $\mathrm{E}(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1; \boldsymbol{\delta}) = \eta(\mathbf{Z}_i, \mathbf{x}_i; \boldsymbol{\delta})$ for a known function $\eta(.)$ of a known infinite-dimensional nuisance parameter $\boldsymbol{\delta}$. Then the previous Appendix shows for each $\boldsymbol{\delta}$ how to calculate the orthogonal complement

of the nuisance tangent space (Bickel, Klaassen, Ritov and Wellner (1993) and Newey (1990)) for $\boldsymbol{\psi}$, which identifies all unbiased estimating functions for $\boldsymbol{\psi}$ (that lead to a RAL estimator). To calculate the orthogonal complement to the nuisance tangent space for $\boldsymbol{\psi}$ in $\mathcal{A}$ defined solely by (8) and the assumptions of Section 2 with $\text{pr}(R_i = 1|\mathbf{x}_i)$ known, we derive the subset of these estimating functions (obtained for all $\boldsymbol{\delta}$) which are orthogonal to the tangent space Λ_δ for the infinite-dimensional nuisance parameter $\boldsymbol{\delta}$. Omitting subscripts i for notational convenience, we first show that the tangent space for $\boldsymbol{\delta}$ in $\mathcal{A}$ is given by

$$\begin{aligned}\Lambda_\delta = \{&Ra_1(Y,\mathbf{Z},\mathbf{x}) + (1-R)a_0(Y_0,\mathbf{x}) : a_1(.), a_0(.) \text{ arbitrary, satisfying}\\ &\text{E}\{a_1(Y,\mathbf{Z},\mathbf{x})|\mathbf{Z},\mathbf{x},R=1\} = 0, \text{E}\{a_0(Y_0,\mathbf{x})|\mathbf{x}\} = 0,\\ &\text{E}\{\epsilon a_1(Y,\mathbf{Z},\mathbf{x})|\mathbf{Z},\mathbf{x},R=1\} = \eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta})(1+\boldsymbol{W}_2'\boldsymbol{\psi}_2),\\ &\text{E}\{Y_0 a_0(Y_0,\mathbf{x})|\mathbf{x}\} = \text{E}\{\eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta})|\mathbf{x},R=1\}\},\end{aligned}$$

where $\epsilon = Y - \boldsymbol{W}_1'\boldsymbol{\psi}_1 - (1+\boldsymbol{W}_2'\boldsymbol{\psi}_2)\eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta})$. To show this, we consider the likelihood for a single observation

$$\begin{aligned}\mathcal{L}(\boldsymbol{\psi},\boldsymbol{\delta},\eta_1,\eta_2,\eta_3,\eta_4) = \eta_4(\mathbf{x})\{&\eta_1(Y|\mathbf{Z},\mathbf{x},R{=}1;\boldsymbol{\psi},\boldsymbol{\delta})\eta_2(\mathbf{Z}|\mathbf{x},R{=}1)\text{pr}(R{=}1|\mathbf{x})\}^R\\ &\times\left\{\int \eta_3(Y_0|\mathbf{Z},\mathbf{x},R{=}1;\boldsymbol{\delta})\eta_2(\mathbf{Z}|\mathbf{x},R{=}1)d\mathbf{Z}\text{pr}(R{=}0|\mathbf{x})\right\}^{1-R},\end{aligned}$$

where $\eta_2(.)$ is the unknown conditional density of $\mathbf{Z}$ (given $(\mathbf{x}, R=1)$), $\eta_1(.), \eta_3(.)$ specify the conditional densities of Y and Y_0 (given $(\mathbf{Z},\mathbf{x},R=1)$) up to the known mean and $\eta_4(.)$ specifies the density of $\mathbf{x}$. It is easy to verify that all scores for $\boldsymbol{\delta}$ are contained in Λ_δ. Vice versa, one can show that all elements of Λ_δ can be viewed as scores for a parametric submodel. Indeed, consider arbitrary $a_1(.), a_0(.)$ satisfying the restrictions for Λ_δ. Next consider the parametric submodel defined by $\mathcal{L}$ with $\eta_1(Y|\mathbf{Z},\mathbf{x},R=1;\delta) = \eta_{10}(Y|\mathbf{Z},\mathbf{x},R=1)\{1+\delta a_1(Y,\mathbf{Z},\mathbf{x})\}$, $\eta_2(\mathbf{Z}|\mathbf{x},R=1) = \eta_{20}(\mathbf{Z}|\mathbf{x},R=1)$, $\eta_3(Y_0|\mathbf{Z},\mathbf{x},R=1;\delta) = \eta_{30}(Y_0|\mathbf{Z},\mathbf{x},R=1)\{1+\delta a_0(Y_0,\mathbf{x})\}$ and $\eta_4(\mathbf{x}) = \eta_{40}(\mathbf{x})$, where $\eta_{10}(.), \eta_{20}(.)$, $\eta_{30}(.)$ and $\eta_{40}(.)$ denote the densities indexing the true regular parametric submodel. Define $\eta_0(\mathbf{Z},\mathbf{x}) = \int Y_0\eta_{30}(Y_0|\mathbf{Z},\mathbf{x},R=1)dY_0$. If the above models are well-defined, it is obvious that the score for δ in these submodels is $Ra_1(Y,\mathbf{Z},\mathbf{x}) + (1-R)a_0(Y_0,\mathbf{x})$ at the truth $\delta = 0$. That these models are well-defined is seen as follows.

1. The density functions are valid in the sense that they integrate to 1, because $\text{E}\{a_1(Y,\mathbf{Z},\mathbf{x})|\mathbf{Z},\mathbf{x},R=1\} = \text{E}\{a_0(Y_0,\mathbf{x})|\mathbf{x}\} = 0$.
2. These parametric submodels satisfy, for all δ, the original restrictions on the observed data law implied by (8). This follows because, for $\eta^*(\mathbf{Z},\mathbf{x}) = \eta_0(\mathbf{Z},\mathbf{x})(1+\delta)$,

$$\begin{aligned}\text{E}(Y|\mathbf{Z},\mathbf{x},R=1) &= \boldsymbol{W}_1'\boldsymbol{\psi}_1 + (1+\boldsymbol{W}_2'\boldsymbol{\psi}_2)\eta_0(\mathbf{Z},\mathbf{x}) + \delta\eta_0(\mathbf{Z},\mathbf{x})(1+\boldsymbol{W}_2'\boldsymbol{\psi}_2)\\ &= \boldsymbol{W}_1'\boldsymbol{\psi}_1 + (1+\boldsymbol{W}_2'\boldsymbol{\psi}_2)\eta^*(\mathbf{Z},\mathbf{x})\end{aligned}$$

by the fact that $\mathrm{E}\{\epsilon a_1(Y,\mathbf{Z},\mathbf{x})|\mathbf{Z},\mathbf{x},R=1\}=\eta_0(\mathbf{Z},\mathbf{x})(1+\boldsymbol{W}_2'\boldsymbol{\psi}_2)$, and

$$\begin{aligned}\mathrm{E}(Y_0|\mathbf{x}) &= \mathrm{E}\{\eta_0(\mathbf{Z},\mathbf{x})|\mathbf{x},R=1\}+\delta\mathrm{E}\{\eta_0(\mathbf{Z},\mathbf{x})|\mathbf{x},R=1\}\\ &= \mathrm{E}\{\eta^*(\mathbf{Z},\mathbf{x})|\mathbf{x},R=1\}\end{aligned}$$

by $\mathrm{E}\{Y_0a_0(Y_0,\mathbf{x})|\mathbf{x}\}=\mathrm{E}\{\eta_0(\mathbf{Z},\mathbf{x})|\mathbf{x},R=1\}$.

Now an arbitrary element $RA_1(Y,\mathbf{Z},\mathbf{x})+(1-R)A_0(Y_0,\mathbf{x})$ of the orthogonal complement to the nuisance tangent space for $\boldsymbol{\psi}$ is orthogonal to the tangent space for δ if $\mathrm{E}\{RA_1(Y,\mathbf{Z},\mathbf{x})a_1(Y,\mathbf{Z},\mathbf{x})+(1-R)A_0(Y_0,\mathbf{x})a_0(Y_0,\mathbf{x})\}=0$. From the previous Appendix, we know that $A_1(Y,\mathbf{Z},\mathbf{x})=\{1-\pi(\mathbf{x})\}[\mathbf{d}_1(\mathbf{x})\{\epsilon+\eta(\mathbf{Z},\mathbf{x})\}+\mathbf{d}_3(\mathbf{x})]+\mathbf{d}_2(\mathbf{Z},\mathbf{x})\epsilon$ for given $\eta(.)$, and that $A_0(Y_0,\mathbf{x})=-\pi(\mathbf{x})\{\mathbf{d}_1(\mathbf{x})Y_0+\mathbf{d}_3(\mathbf{x})\}$. Calculating the above expectation and solving the equation yields

$$\mathbf{d}_2(\mathbf{Z},\mathbf{x})=-\{1-\pi(\mathbf{x})\}\mathbf{d}_1(\mathbf{x})\frac{\boldsymbol{W}_2'\boldsymbol{\psi}_2}{1+\boldsymbol{W}_2'\boldsymbol{\psi}_2}.$$

From this, we find that the orthogonal complement of the nuisance tangent space for $\boldsymbol{\psi}$ in $\mathcal{A}$ is

$$\Big\{\{R-\pi(\mathbf{x})\}\Big[\mathbf{d}_1^*(\mathbf{x})\Big(R\frac{Y-\boldsymbol{W}_1'\boldsymbol{\psi}_1}{1+\boldsymbol{W}_2'\boldsymbol{\psi}_2}+(1-R)Y\Big)+\mathbf{d}_2^*(\mathbf{x})\Big];\mathbf{d}_1^*(.),\mathbf{d}_2^*(.)\text{arbitrary}\Big\}.$$

To find the efficient score, we project the score for $\boldsymbol{\psi}$, i.e.,

$$-R\frac{\partial\log\eta_1(Y,\mathbf{Z},\mathbf{x};\boldsymbol{\psi},\boldsymbol{\delta})}{\partial\epsilon}\left(\boldsymbol{W}_1'\ \boldsymbol{W}_2'\eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta})\right),$$

with ϵ defined earlier and $\eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta})=\boldsymbol{W}_3'\boldsymbol{\delta}$, onto the orthogonal complement of the nuisance tangent space for $\boldsymbol{\psi}$ in $\mathcal{A}$. It can easily be checked that this projection is obtained by setting

$$\begin{aligned}\mathbf{d}_2(\mathbf{x}) &= -\mathbf{d}_1(\mathbf{x})\mathrm{E}(H|\mathbf{x}),\\ \mathbf{d}_1(\mathbf{x}) &= \frac{\mathrm{E}\Big\{\frac{(\boldsymbol{W}_1'\ \boldsymbol{W}_2'\eta(\mathbf{Z},\mathbf{x};\boldsymbol{\delta}))}{1+\boldsymbol{W}_2'\boldsymbol{\psi}_2}|\mathbf{x},R=1\Big\}}{\{1-\pi(\mathbf{x})\}\mathrm{Var}\,(H|\mathbf{x},R=1)+\pi(\mathbf{x})\mathrm{Var}\,(H|\mathbf{x},R=0)}.\end{aligned}$$

A.3. Proof of nonparametric identifiability

Denote expectations and densities w.r.t. the true observed data law with subscript 0, and those based on the model without subscript. To prove the theorem, we construct a joint law of $(Y_i,Y_{i0},\mathbf{Z}_i,R_i,\mathbf{x}_i)$ that marginalizes to the observed data law and satisfies all assumptions. First we set $f(\mathbf{Z}_i,\mathbf{x}_i,R_i)=f_0(\mathbf{Z}_i,\mathbf{x}_i,R_i)$. Then we choose $f(Y_i|\mathbf{Z}_i,\mathbf{x}_i,R_i)=f_0(Y_i|\mathbf{Z}_i,\mathbf{x}_i,R_i)$ and define $Y_{i0}=$

Y_i in the subpopulation defined by $(\mathbf{Z}_i = \mathbf{0}, \mathbf{x}_i, R_i)$ to satisfy (A4). To satisfy the model restrictions, we choose $h_1(\mathbf{Z}_i, \mathbf{x}_i)$ and $h_2(\mathbf{x}_i)$ as the solutions to

$$0 = \mathrm{E}_0\{Y_i - h_2(\mathbf{x}_i)Z_{i1} - q(\mathbf{Z}_i, \mathbf{x}_i)h_1(\mathbf{Z}_i, \mathbf{x}_i)|\mathbf{x}_i, R_i = 1\} - \mathrm{E}_0(Y_i|\mathbf{x}_i, R_i = 0), \quad (19)$$

$$0 = \mathrm{E}_0[Y_i - h_2(\mathbf{x}_i)Z_{i1} - \{1 + q(\mathbf{Z}_i, \mathbf{x}_i)\}h_1(\mathbf{Z}_i, \mathbf{x}_i)|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1]. \quad (20)$$

These solutions are unique when, for each $\mathbf{x}_i$, $\mathrm{E}[Z_{i1}/\{1 + q(\mathbf{Z}_i, \mathbf{x}_i)\} \mid \mathbf{x}_i, R_i = 1]$ differs from 0 and for each $(\mathbf{Z}_i, \mathbf{x}_i)$, $q(\mathbf{Z}_i, \mathbf{x}_i)$ does not equal -1. Finally, let $f(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ be an arbitrary density with conditional mean $h_1(\mathbf{Z}_i, \mathbf{x}_i)$, given $(\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$, that satisfies $f(Y_{i0}|\mathbf{Z}_i = \mathbf{0}, \mathbf{x}_i, R_i = 1) = f_0(Y_i|\mathbf{Z}_i = \mathbf{0}, \mathbf{x}_i, R_i = 1)$ abd (A1). This is possible by (19) and (20). Likewise, let $f(Y_i|Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$ be any density with conditional mean $h_2(\mathbf{x}_i)Z_{i1} + \{1 + q(\mathbf{Z}_i, \mathbf{x}_i)\}Y_{i0}$, given $(Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$, that satisfies $\int f(Y_i|Y_{i0}, \mathbf{Z}_i, \mathbf{x}_i, R_i = 1)f(Y_{i0}|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)dY_{i0} = f_0(Y_i|\mathbf{Z}_i, \mathbf{x}_i, R_i = 1)$. Doing this is possible by (20). We conclude that the unknown functions $h_1(.)$ and $h_2(.)$ are identified and that the chosen full data law satisfies the model restrictions and marginalizes to the observed data law by construction. The lack of identifiability of $q(\mathbf{Z}_i, \mathbf{x}_i)$ follows because, by repeating the earlier arguments for different values of $q(\mathbf{Z}_i, \mathbf{x}_i) \neq -1$, it is seen that all such values are compatible with the observed data law.

References

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *J. Amer. Statist. Assoc.* **91**, 444-455.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer-Verlag, New-York.

Chamberlain, G. (1992). Comment: Sequential moment restrictions in panel data. *J. Bus. Econom. Statist.* **10**, 20-26.

Cuzick, J., Edwards, R. and Segnan, N. (1997). Adjusting for non-compliance and contamination in randomized clinical trials. *Statist. Medicine* **16**, 1017-1029.

Efron, B. and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials (with discussion). *J. Amer. Statist. Assoc.* **86**, 9-26.

Fischer-Lapp, K. and Goetghebeur, E. (1999). Practical properties of some structural mean analyses of the effect of compliance in randomized clinical trials. *Cont. Clin. Trials* **20**, 531-546.

Fischer-Lapp, K. and Goetghebeur, E. Structural mean effect of noncompliance: estimating interaction with baseline prognosis and selection effects. Accepted for publication in *J. Amer. Statist. Assoc.* To appear.

Goetghebeur, E. and Lapp, K. (1997). The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Appl. Statist.* **46**, 351-364.

Goetghebeur, E. and Lapp, K. (1998). Estimating effectiveness and efficacy at different observed dosing levels with selective noncompliance in a placebo-controlled clinical trial. *Proceedings of the* 1998 *Joint Statistical Meetings of the American Statistical Association, the Biometric Society and the Institute of Mathematical Statistics*, 61-70.

Newey, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5**, 99-135.

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Method* **23**, 2379-2412.

Robins, J. M. (1997a). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (Edited by M. Berkane), 69-117. Springer Verlag, New-York.

Robins, J. M. (1997b). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statist. Medicine* **16**, 21-37.

Robins, J. M. and Rotnitzky, A. Noncompliance and instrumental variable estimation of logistic structural models. Accepted for publication in *Biometrika*. To appear.

Rubin, D. B. (1978). Bayesian inference for causal effects, *Ann. Statist.* **6**, 34-58.

Vansteelandt, S. and Goetghebeur, E. (2003). Causal inference with generalized structural mean models. *J. Roy. Statist. Soc. Ser. B* **65**, 817-835.

Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281-S9, 9000 Ghent, Belgium.

E-mail: Stijn.Vansteelandt@UGent.be

Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281-S9, 9000 Ghent, Belgium.

E-mail: Els.Goetghebeur@UGent.be

(Received January 2003; accepted August 2003)