Maximum Likelihood Estimation of Structural Nested Logistic Model with an Instrumental Variable

Roland A. Matsouaka *
Department of Epidemiology
Harvard University
Boston, MA 02115

Eric J. Tchetgen Tchetgen
Departments of Biostatistics and Epidemiology
Harvard University
Boston, MA 02115

Abstract

Current estimating equation methods for logistic structural nested mean models (SNMMs) either rely heavily on possible ”uncongenial” modeling assumptions or involve a cumbersome integral equation needing to be solved for each independent unit at each step of solving the estimating equation. These drawbacks have impeded widespread use of these methods. In this paper, we present an alternative parametrization of the likelihood function for the logistic SNMM that circumvents computational complexity of existing methods while ensuring a congenial parametrization of SNMM. We also provide a goodness-of-fit test for evaluating parametric assumptions made by the likelihood model. Our method can be easily implemented using most standard statistical softwares, and is illustrated via a simulation study.

1 Introduction and Background

Structural nested mean models (SNMMs) and G-estimation were introduced by Robins (1989, 1994) as rigorous statistical approaches to infer causality in studies where exposure (or treatment) assignments are not completely under the control of investigators, such as observational studies or clinical trials with non-compliance.

Inherent to all observational studies and clinical trials with non-compliance is the issue of confounding or non-ignorable selection of the exposure. If this issue is not taken into account, the estimated effects of exposure on the outcome can be biased and inconsistent, leading to spurious results. Instrumental variables (IV) have been used profusely in the literature to estimate the effects of exposure on the outcome when unobserved confounding is present or suspected.

An IV for the effect of exposure X on an outcome Y is a pre-treatment variable Z that, given a set of measured baseline covariates L, is (1) associated with X, but (2) associated with Y only through X (i.e. not direct effect of Z on Y, also known as exclusion restriction), and (3) independent of any unmeasured confounding variable U of the effects of X on Y (see Figure 1).

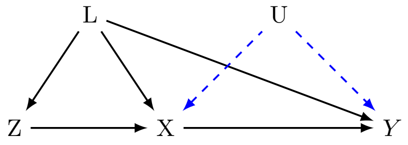

Figure 1: A graph showing an instrumental variable Z, a measured covariate L, an unmeasured confounding U, an exposure X, and an outcome Y

[*] email: rmatsoua@hsph.harvard.edu

The assumptions can also be given in term of potential outcomes (see Neyman (1923), Rubin (1978), or Robins (1986)). Let us define the potential outcome Y_{xz} as the value of the outcome of interest had, possibly contrary to fact, Z been set to z and X set to x by external intervention. Likewise, Y_x denotes the outcome had exposure been set to x. L is a vector of measured baseline covariates; the vector L will typically include all measured confounders of the effects of Z on (X, Y) and for the effects of X on Y. Conditions (1) to (3) can be expressed as: (1) non-null association between X and Z i.e. $X \not\perp Z|L$; (2) exclusion restriction i.e. $Y_{xz} = Y_x$, almost surely for all x, z and (3) independence of potential outcomes and IV i.e. $Y_{xz} \perp Z|L$, for all x, z. The notation $A \perp B|C$ indicates stochastic independence between random variables A and B given C.

From assumptions (2) and (3) we derive a relatively weaker assumption, $E(Y_0|z, l) = E(Y_0|l)$, which plays an important role in parameter identification.

Let $W = (Z, L)$, we define SNMMs as

$$b(E(Y_x|x, w)) - b(E(Y_0|x, w) = \gamma(x, w),$$

where b is the link function. The function b is the identity, the log, and the logit function for, respectively, the additive, multiplicative, and logistic SNMMs. The contrast γ compares the average potential outcomes under active and inactive treatment values on a scale given by b, for the subset of the population with (x, w). Therefore, γ constitutes a conditional causal effect.

Robins (1994) establishes that assumptions (1) to (3) do not suffice to nonparametrically identify γ in the case of the identity link. Robins and Rotnitzky (2004) (hereafter RR) further show that assumption (1) to (3) are likewise insufficient for identification for the logit link (see also Richardson and Robins (2010) or Didelez, Meng and Sheehan (2010)). Thus, to proceed, we require an additional assumptions for identification (see RR): (4) $\gamma(x, w)$ is restricted such that it is identified. We let *psi* index such a model $\gamma(x, w) = \gamma(x, w; \psi)$. In the event that X and Z are binary, Robins (1994) takes as assumptions the "no current treatment value interaction" assumption: $\gamma(x, z, l) = \gamma(x, l)$. He shows that the function $\gamma(x, w; \psi)$ is completely identify in this case.

It is worth mentioning that the assumptions (4) and (5) are not empirically verifiable. Vansteelandt and Goetghebeur (2005) study the violation of assumption (4). Alternative identification conditions other than (4) and (5) are of great interest and constitute an important resarch topic. Richardson and Robins (2010) and Richardson, Evans, and Robins (2011) elucidate the issue of identification and provide studies of the boundaries of the IV models. Here we assume that the parametric model $\gamma(x, w; \psi)$ is correctly specified.

Treatment effects for additive and multiplicative SNMMs can be estimated via G-estimation. As shown by Robins, G-estimation is advantageous as it avoids having to estimate the baseline mean $E(Y_0|x, w)$ in order to estimate γ, provided one can correctly specify a model for the conditional density of Z given L. In randomized experiments, this density is known by design, in which case G-estimation is guaranteed to be consistent, in particular under the null hypothesis. In general, G-estimators for these models are consistent, asymptotically normal, and (can be) semi-parametrically efficient, assuming at least correct model for Z (Robins (1989, 1994) and RR). As we said previously, Robins et al. (1999) and RR show that logistic SNMMs cannot be estimated with G-estimation. Specifically, they show that it is not possible to construct an estimator of γ that is regular and asymptotically linear when the density of Z is known, without also needing to estimate the baseline mean $E(Y_0|x, w)$.

Vansteenlandt and Goetghebeur (2003) (hereafter VG) propose an "association" model

$$\text{logit}\{E(Y|x, w)\} = m(x, w; \eta),$$

which they use together with a model for the IV to estimate γ. However, when the association model is not saturated, it can be uncongenial (in the sense described by Meng (1994)) to the logistic SNMM model i.e. the two models can be incompatible or inconsistent (see RR, VG, or Vansteenlandt et al. (2011)).

As an alternative, RR propose a different parametrization based on the contrast

$$\text{logit}E(Y_0|x,w) - \text{logit}E(Y_0|x=0,w) = q(x,w;\eta),$$

that is always guaranteed to be congenial. This parametrization implies that

$$P(Y=1|x,w) = \gamma(x,w) + q(x,w) + \\ b(E(Y_0|x=0,w)),$$

where $v(w) = b(E(Y_0|x=0,w))$ is the unique solution to the integral equation

$$\text{logit} \int \text{expit}\{q(x^*,w) + v(w)\} dF(x^*|w)dx \\ = \text{logit}\{E(Y_0|l)\} = t(l). \quad (1)$$

This integral equation cannot be solved for v, in closed form—for most choices of models for $q(x,w)$, $f(x|w)$ and $t(l)$—except say when x is binary (see RR or Vansteenlandt et al. (2011)). Thus, a numerical optimization of the parametric likelihood for the joint density of the observables using the parametrization proposed by RR involves solving numerically an integral equation for each observed $W=(Z,L)$, within each iteration of the algorithm. Unfortunately, when the exposure takes more than two values, or is continuous or multivariate, RR approach becomes computationally challenging, particularly when the IV is continuous and there are a large number of covariates L. This numerical drawback has impeded the widespread use of this approach, despite its mathematical and theoretical underpinning.

The purpose of this paper is to describe an alternative strategy for estimating the parameters of a logistic SNMM under assumptions (1) to (3) using a novel parametrization. In order to estimate the logistic SNMM, we propose a likelihood approach, and give a goodness-of-fit test, that builds directly on the work of VG and RR. The novelty of our approach is that, unlike VG and similar to RR, we use a variation independent congenial parametrization of the observed data likelihood. However, unlike RR, our approach does not involve solving integral equations and is, therefore, readily implementable regardless of the nature or the dimension of the exposure, IV, and covariates.

In Section 2, we introduce the new parametrization and relate it to VG and RR, respectively. In Section 3, we present some important properties of the proposed parametrization and provide a goodness-of-fit test for evaluating the parametric assumptions of the fitted likelihood model. Then, in Section 4, we run simulation studies for both binary and continuous exposure using continuous baseline covariates. We also closely examine the performance of the goodness-of-fit test. Finally, in Section 5, we close with some final remarks.

2 New Parametrization

Using the notation of RR, suppose we observe n independent and identically distributed copies of the vector $O=(L,X,Z,Y)$ and we wish to estimate the parameter ψ of the logistic SNMM

$$\text{logit}P(Y_x=1|x,w) - \text{logit}P(Y_0=1|x,w) \\ = \gamma(x,w;\psi) \quad (2)$$

under the assumption that Z is a valid IV.

The observed data likelihood factorizes as

$$f_y(Y|X,W;\psi)f_x(X,f_z(Z|L)f_l(L) \\ = f_y(Y_X|X,W;\psi)f_x(X,f_z(Z|L)f_l(L).$$

We write f_y as

$$\text{logit}f_y(1|x,w) = \text{logit}P(Y_x=1|x,w) \\ = \text{logit}P(Y_x=1|x,w) - \text{logit}P(Y_0=1|x,w) \\ + \text{logit}P(Y_0=1|x,w) - \text{logit}P(Y_0=1|x=0,w)$$

$$
\begin{aligned}
&+ \operatorname{logit} P(Y_0 = 1|x = 0, w) \\
&= \gamma(x, w; \psi) + q(x, w) + v(w) \qquad (3)
\end{aligned}
$$

where $v(w) =$ logit$P(Y_0 = 1|x = 0, w)$ and

$$
\begin{aligned}
q(x, w) = {} & \operatorname{logit} P(Y_0 = 1|x, w) \\
& - \operatorname{logit} P(Y_0 = 1|x = 0, w).
\end{aligned}
$$

The function q encodes the degree of unobserved confounding and is sometimes referred to a selection bias function. As RR point out, γ and q are variation independent, but v is not a free parameter as it must satisfy the restriction

$$
\begin{aligned}
\int P(Y_0 = 1|x^*, w) dF_x(x^*|w) &= P(Y_0 = 1|w). \\
&= P(Y_0 = 1|l) \qquad (4)
\end{aligned}
$$

Let $b = \Phi^{-1}$ denote the logit link, let $f_{x,0}$ and f_x be the densities of $F_{x,0}$ and F_x with respect to a dominating measure μ, and consider $t(l) = \Phi^{-1}P(Y_0 = 1|l)$, equation (4) becomes

$$
\int \Phi(q(x^*, w) + v(w)) dF_x(x^*|w) = \Phi(t(l)).
$$

Thus, v is a functional of q, F_x and t implicitly defined by the integral equation (4) that must be solved for each observation. Note that, as mentioned by RR and unlike VG, equation (4) guarantees congeniality.

Consider parametric models $q(x, w; \eta)$, $t(l; \omega)$, and $F_x(x|w; \alpha)$, the resulting observed data likelihood is given by

$$
f_y(Y|X, W; \psi, \eta, \alpha, \omega) f_x(X|W; \alpha) f_z(Z|L) f_l(L)
$$

with logit$f_y(1|x, w; \psi, \eta, \alpha, \omega)$ satisfying (3). This is the congenial parametrization of RR.

We now propose an alternative congenial parametrization that obviates the need to solve integral equation (4). To proceed, note that $v(w) =$ logit$P(Y_0 = 1|x = 0, w) =$ $-$ [logit$P(Y_0 = 1|w) -$ logit$P(Y_0 = 1|x = 0, w)] +$ logit$P(Y_0 = 1|w)$. Further, recall that

$$
\operatorname{logit} P(Y_0 = 1|w) = \log \frac{P(Y_0 = 1|w)}{P(Y_0 = 0|w)} = \log \operatorname{ODDS}(w).
$$

Define

$$
\overline{q}(w) = \log \int \exp[q(x, w; \eta)] dF_{x,0}(x|w, Y_0 = 0)
$$

Since

$$
\begin{aligned}
\operatorname{ODDS}(w) &= \int \operatorname{ODDS}(x, w) dF_{x,0}(x|w, Y_0 = 0) \\
&= \int \frac{P(Y_0 = 1|x, w)}{P(Y_0 = 0|x, w)} dF_{x,0}(x|w, Y_0 = 0),
\end{aligned}
$$

with $F_{x,0}(x|w, Y_0 = 0)$ the CDF of X given W and $Y_0 = 0$, it follows that

$$
\begin{aligned}
& v(w) = - [\operatorname{logit} P(Y_0 = 1|w) \\
& - \operatorname{logit} P(Y_0 = 1|x = 0, w)] + \operatorname{logit} P(Y_0 = 1|w) \\
& = - \log \left[\frac{P(Y_0 = 1|w)}{1 - P(Y_0 = 1|w)} \right] \\
& \qquad - \log \left[\frac{1 - P(Y_0 = 1|x = 0, w)}{P(Y_0 = 1|x = 0, w)} \right] + t(l) \\
& = - \log \int \operatorname{ODDS}(x, w) dF_{x,0}(x|w, Y_0 = 0) \\
& \qquad + \log [\operatorname{ODDS}(x = 0, w)] + t(l) \\
& = - \log \int \frac{\operatorname{ODDS}(x, w)}{\operatorname{ODDS}(x = 0, w)} dF_{x,0}(x|w, Y_0 = 0) + t(l) \\
& = - \log \int \exp[q(x, w; \eta)] dF_{x,0}(x|w, Y_0 = 0) + t(l) \\
& = -\overline{q}(w) + t(l).
\end{aligned}
$$

$$
\begin{aligned}
\text{Thus, } \operatorname{logit} f_y(1|x, w) &= \gamma(x, w; \psi) + q(x, w) \\
&\quad - \overline{q}(w) + t(l). \qquad (5)
\end{aligned}
$$

This means that in this new parametrization, we are free to choose models for $q(x, w)$, $f_{x,0}(x|w, Y_0 = 0)$, and $t(l)$. However, the density $f_x(x|w)$ is fully determined by the above parameters according to the following expression

$$
\begin{aligned}
f_x(x|w) &= f_{x,0}(x|w, Y_0 = 0) P(Y_0 = 0|w) + \\
&\qquad f_{x,0}(x|w, Y_0 = 1) P(Y_0 = 1|w) \\
&= f_{x,0}(x|w, Y_0 = 0) \; (1 - \Phi(t(l))) + \\
&\quad \frac{f_{x,0}(x|w, Y_0 = 0) \exp(q(x, w))}{\int \exp(q(x^*, w)) dF_{x,0}(x^*|w, Y_0 = 0)} \Phi(t(l))
\end{aligned}
$$

i.e. $$f_x(x|w) = f_{x,0}(x|w, Y_0 = 0)\ (1 - \Phi(t(l)))+ \frac{\exp(q(x, w))}{\exp(\overline{q}(w))} f_{x,0}(x|w, Y_0 = 0)\Phi(t(l)) \qquad (6)$$

3 Characteristics of the New Parametrization

In this section, we present some key results related to the proposed parametrization.

3.1 Integral Property

Theorem 1: Under assumptions (1) and (3) and the proposed parametrization, we have

$$\int P(Y_0 = 1|x^*, w)dF_x(x^*|w) = P(Y_0 = 1|w)$$
$$= P(Y_0 = 1|l) \qquad \Leftrightarrow$$
$$\int \Phi(q(x^*, w) - \overline{q}(w) + t(l))dF_x(x^*|w) = \Phi(t(l))$$

Proof: Consider the following representation of the joint density of Y_0 and X given W, (see for example, Tchetgen Tchetgen, Robins, and Rotnitzky (2010))

$$f(Y_0 = y, X = x^*|w) = D_0^{-1}(w)P(Y_0 = y|X = 0, w) \times \mathrm{OR}_0(x^*|w)^y f_{x,0}(x^*|Y_0 = 0, w),$$

with

$$D_0(w) = \int \sum_{y^*} P(Y_0 = y^*|x = 0, w)\mathrm{OR}_0(x^*|w)^{y^*} \times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*)$$
$$= \frac{P(Y_0 = 0|X = 0, w)}{P(Y_0 = 0|w)}, \quad \text{and}$$
$$\mathrm{OR}_0(x|w) = \frac{P(Y_0 = 1|x, w)P(Y_0 = 0|x = 0, w)}{P(Y_0 = 0|x, w)P(Y_0 = 1|x = 0, w)}.$$

We have

$$\int P(Y_0 = 1|x^*, w)dF_x(x^*|w) = \int P(Y_0 = 1, x^*|w)d\mu(x^*)$$
$$= D_0^{-1}(w) \int P(Y_0 = 1|x = 0, w)\mathrm{OR}_0(x^*|w)\times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*).$$

Thus,

$$\int P(Y_0 = 1|x^*, w)dF_x(x^*|w)$$
$$= D_1^{-1}(w)\int \frac{P(Y_0 = 1|x = 0, w)}{P(Y_0 = 0|x = 0, w)}\mathrm{OR}_0(x^*|w)\times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*)$$
$$= D_1^{-1}(w)\int \exp(q(x^*, w) - \overline{q}(w) + t(l))\times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*) = \Phi(t(l))$$

where

$$D_1(w) = \sum_y \int \frac{P(Y_0 = y|x = 0, w)}{P(Y_0 = 0|x = 0, w)}\mathrm{OR}_0(x^*|w)^y \times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*)$$
$$= 1 + \int \exp(q(x^*, w) - \overline{q}(w) + t(l))\times f_{x,0}(x^*|Y_0 = 0, w)d\mu(x^*)$$

Q.E.D.

3.2 Parameter Estimation

3.2.1 Maximum likelihood estimation

We construct the MLE of ψ using parametric models

(1) $f_{x,0}(x|Y_0 = 0, w; \alpha)$

(2) $\mathrm{logit} f_y(1|x, w; \psi_0, \eta, \alpha, \omega) = \gamma(x, w; \psi) + q(x, w; \eta) - \overline{q}(h; \eta, \alpha) + t(l, \omega);$

which gives

$$f_x(x|w; \alpha, \omega, \eta) = f_{x,0}(x|w, Y_0 = 0; \alpha)(1 - \Phi(t(l; \omega))) + \exp(q(x^*, w) - \overline{q}(w))f_{x,0}(x|w, Y_0 = 0; \alpha))\Phi(t(l; \omega)).$$

The resulting likelihood is given by:

$$\prod f_y(Y_i|X_i, Z_i, L_i; \psi, \eta, \alpha, \omega)f_x(X_i|Z_i, L_i; \alpha, \omega, \eta)\times f_z(Z|L; \kappa)$$

The above likelihood can be maximized using PROC NLMIXED in SAS or the *optim* function in R. As opposed to the integral equation (1), the integral

$$\overline{q}(w) = \int \exp[q(x, w; \eta)] dF_{x,0}(x|w, Y_0 = 0) + t(l)$$

is easy to implement numerically even when there is no close for representation. In that case, one can use Gauss-Hermite quadrature integral approximation.

Note that the MLE of $(\psi, \eta, \alpha, \omega)$ is uncorrelated with that of κ. In fact, we need not estimate the latter to obtain an estimate of the former. This, in turn, implies that the MLE of ψ cannot exploit any prior information about f_z such as the known randomization probability in a randomized experiment. This is a notable limitation of the likelihood approach. To remedy this problem VG and RR propose methods that are doubly robust under the sharp null hypothesis, $\gamma = 0$, of no exposure causal effect by explicitly using any knowledge about f_z. RR, in particular, propose to use an influence function function of ψ for inference, which is endowed with the above robustness property, but suffers the same computational limitations as their likelihood approach. An alternative approach to the likelihood will be to solve RR estimating equation using our parametrization.

3.2.2 Estimating Equation Based Goodness-of-fit Test

RR characterize the class of influence functions (see Tsiatis (2006) for a definition) for γ in the semiparametric model where γ is assumed to follow a parametric specification and the likelihood is otherwise unrestricted. We use a scalar function from this class to construct a semiparametric goodness-of-fit test for the likelihood model. Let

$$\begin{aligned}\widehat{M_1} &= \gamma(X, W; \widehat{\psi}) + q(X, W; \widehat{\eta}) \\ &\quad - \overline{q}(W; \widehat{\eta}, \widehat{\alpha}) + t(L, \widehat{\omega}), \\ \widehat{M_2} &= q(X, W; \widehat{\eta}) - \overline{q}(W; \widehat{\eta}, \widehat{\alpha}) + t(L, \widehat{\omega})\end{aligned}$$

where $(\widehat{\psi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\omega})$ is the MLE obtained in the previous section. Consider the goodness-of-fit statistic $\widehat{U} = U(\widehat{\psi}, \widehat{\eta}, \widehat{\alpha}, \widehat{\omega}) = (Z - \mathbb{E}(Z|L; \widehat{\kappa}))\times$

$$\left[\frac{\widehat{\Phi}(M_2)(1-\widehat{\Phi}(M_2))}{\widehat{\Phi}(M_1)(1-\widehat{\Phi}(M_1))}(Y-\Phi(\widehat{M_1}))+\Phi(\widehat{M_2})-\Phi(t(L,\widehat{\omega}))\right].$$

Theorem 2: Under the null hypothesis that the likelihood model is correclty specify, $\Phi(\widehat{M_1})$ converges in probability to $P(Y = 1|X, W)$ and we have

$$\Omega \leq \mathbb{E}(U(\psi, \eta, \alpha, \omega)^2)$$

with Ω the asymptotic variance of $n^{1/2} \sum_i \widehat{U_i}$.

Therefore, under H_0^*, $P(|T| > 1.96)$ converges with increasing sample size to $c \leq 0.05$, where

$$T = \frac{\sum_i \widehat{U_i}}{\sqrt{\sum_i \widehat{U_i}^2}} \tag{7}$$

Proof: RR has established that U is an influence function for γ in the semiparametric model where only γ is parametric, and the rest of the model is nonparametric. This, in turn, implies that $\mathbb{E}\left[\frac{\partial U(\psi_0,\eta,\alpha,\omega)}{\partial(\eta,\alpha,\omega)}|_{(\eta,\alpha,\omega)=(\eta_0,\alpha_0,\omega_0)}\right] = 0$. Hence, under the null hypothesis, $\displaystyle\sum_i \widehat{U}_i \approx$

$$\sum_i \left\{U_i + \mathbb{E}\left[\frac{\partial U(\psi)}{\partial \psi}|_{\psi=\psi_0}\right]\mathbb{E}\left[S_\psi^{\mathit{eff}} S_\psi^{\mathit{eff}T}\right]^{-1} S_{\psi,i}^{\mathit{eff}}\right\}$$

where S_ψ^{eff} is the efficient score of ψ under our parametric model. By a property of influence functions, $\mathbb{E}\left[\frac{\partial U(\psi)}{\partial\psi}|_{\psi=\psi_0}\right] = -\mathbb{E}\left[U S_\psi^{\mathit{eff}}\right]$.

Thus, $\displaystyle n^{-1/2}\sum_i \widehat{U} \approx n^{-1/2}\sum_i \left\{U_i - \mathbb{E}\left[U S_\psi^{\mathit{eff}}\right]\right.$ $\left.\times\mathbb{E}\left[S_\psi^{\mathit{eff}} S_\psi^{\mathit{eff}T}\right]^{-1} S_{\psi,i}^{\mathit{eff}}\right\}$ and we have

$$\begin{aligned}\mathbb{E}\left[\left\{U_i - \mathbb{E}\left[U S_\psi^{\mathit{eff}T}\right]\mathbb{E}\left[S_\psi^{\mathit{eff}} S_\psi^{\mathit{eff}T}\right]^{-1} S_{\psi,i}^{\mathit{eff}}\right\}^2\right] \\ \leq \mathbb{E}\left\{U_i^2\right\}\end{aligned} \tag{8}$$

since $\mathbb{E}\left[U S_\psi^{\mathit{eff}T}\right]\mathbb{E}\left[S_\psi^{\mathit{eff}} S_\psi^{\mathit{eff}T}\right]^{-1} S_{\psi,i}^{\mathit{eff}}$ is the orthogonal projection of U onto the span of $S_{\psi,i}^{\mathit{eff}}$, proving the first result. The second result follows from a standard application of Slutsky's theorem and the central limit theorem.

Q.E.D.

This is a useful result because it states that, assuming our model for γ and for the density of Z given L is correctly specified, T is a valid test statistic of the null hypothesis that the likelihood model is correctly specified. Furthermore, the theorem shows that the test statistic is easily computed without the need to compute the exact variance of $\widehat{U}$, which may be computationally demanding.

The advantage of using an influence function to construct the GOF is that under the null hypothesis that the likelihood model is correct, the GOF statistic already accounts for the variability associated with estimation of all nuisance parameters. The theorem states, through equation (8), that further ignoring the variability due to $\widehat{\psi}$ results in conservative GOF test. In addition, the test is expected to be consistent—against the alternative where we considered γ is correct and only test the remainder of the likelihood—since if the model is mis-specified, we expect that $\mathbb{E}(U) \neq 0$.

4 Simulation Study

In this section, we provide an algorithm to generate data (W, X, Y) following our proposed parametrization, for binary and continuous exposure.

For our simulations, we sample $L = (L_1, L_2)$ from independent bivariate normal or Bernoulli, then generate Z binary from a logistic regression $f_z(Z|L;\kappa)$. Let $\Phi(t(L,\omega))$ be a standard logistic regression with parameter ω and $f_{x,0}(x|w, Y_0 = 0; \alpha)$ a logistic regression with parameter α if X is binary or a normal density function if X is continuous. Let $q(X, W; \eta)$ a simple model for the selection bias function, say $q(X, W; \eta) = \eta X$. We generate X from the density $f_x(X_i|W_i; \alpha, \omega, \eta)$ defined in (6) and Y from the logistic regression model with event probability $\Phi(\gamma(x, w; \psi)+q(x, w; \eta)-\bar{q}(h; \eta, \alpha)+t(l, \omega))$, with a simple choice for the parametric model $\gamma(x, w; \psi_0)$, say $\gamma(x, w; \psi) = \gamma x$.

More precisely, we generated $L_1{\sim}N(3, 1)$, $L_2{\sim}N(2, 1)$; $Z{\sim}Bernoulli(p_z)$, $\text{logit}(p_z){=}\kappa_0{+}\ \kappa_1 L_1{+}\kappa_2 L_2 = -0.1{+}0.5L_1{+}0.2L_2$; and $t(L){=}\ \omega_0{+}\omega_1 L_1{+}\omega_2 L_2{=}{-1}{+}0.5L_1{+}0.3L_2$. For $q(X, W){=}\eta X{=}{-0.4}X$, we derive the marginal distributions of X using (6). Finally, we generated the outcome such that $Y{\sim}Bernoulli(p_y)$, with $\text{logit}(p_y) = (1 - 0.4)X - \bar{q}(W) + t(L)$.

Overall, we generated a total of 2000 data sets of size $n = 5000$ and estimated the model parameters, the empirical type I error, and the power of the goodness-of-fit (GOF) test at 5% significant level i.e. the proportion of simulated data sets for which $|T| > 1.96$. Models fitting was performed using PROC NLMIXED in SAS.

4.1 Binary Exposure

Consider $X|W,Y_0 = 0 \sim Bernoulli(p_{x0})$ with $\text{logit}(p_{x0}) = -0.4{-}0.3L_1{+}0.3L_2 + Z$.
Let $p_t = \Phi(t(L))$. Using equations (6), one can show that $\bar{q}(W) = \log\left[\dfrac{\exp(\text{logit}(p_{x1}))}{1 + \exp(\text{logit}(p_{x0}))}\right]$ and $X{\sim}Bernoulli(p_x)$, $p_x = (1 - p_t) \times p_{x0} + p_t \times p_{x1}$ where $\text{logit}(p_{x1}) = \text{logit}(p_{x0}) - 0.4$.

Table 1: Estimation Results: Binary Exposure

Parameter	Bias	MSE	Coverage	S.E.
ψ	0.002	0.102	0.96	0.319
η	-0.004	0.106	0.95	0.326
α_0	0.005	0.007	0.95	0.083
α_1	-0.002	0.001	0.95	0.038
α_2	-0.001	0.001	0.94	0.033
α_3	0.001	0.003	0.95	0.054
ω_0	0.006	0.031	0.96	0.177
ω_1	0.000	0.002	0.95	0.041
ω_2	-0.001	0.002	0.95	0.040
κ_0	-0.001	0.004	0.95	0.065
κ_1	0.000	0.001	0.94	0.032
κ_2	0.001	0.000	0.95	0.030

Corresponding GOF test: Type I error = 0.01

In Table 1, we present estimation results for all parameters and the goodness-of-fit (GOF) test type I error. Estimation results show a good performance of the likelihood method under our proposed parametrization, with small bias and good coverage probability. The goodness-of-fit test rejects the null hypothesis less often and thus has low power, which is not surprising since we opted for a conservative variance estimate of the goodness-of-fit statistic.

4.2 Continuous Exposure

Consider $X|W,Y_0 = 0 \sim N(\mu_{x0}, \sigma^2)$ with $\mu_{x0}{=}1{-}2L_1{+}L_2{+}3Z$. We have $\bar{q}(W){=}{-}0.4(\mu_{x0} - 0.2\sigma^2)$ and show that X follows a mixture of two normal distributions with density $f(x) = (1 - p_t)f_{0x}(x) + p_t f_{1x}(x)$ where, $f_{kx}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu_{0x} + 0.4k\sigma^2)^2}{2\sigma^2}\right]$.

Estimation results and the performance of the GOF test are summarized in Table 2. Similar to binary exposure, the MLE appears to perform very well, but the GOF is quite conservative.

Table 2: Estimation Results: Cont. Exposure

Parameter	Bias	MSE	Coverage	S.E.
ψ	0.012	0.001	0.95	0.042
η	0.000	0.003	0.95	0.052
α_0	0.000	0.004	0.94	0.062
α_1	0.000	0.000	0.95	0.016
α_2	0.000	0.000	0.95	0.014
α_3	-0.001	0.002	0.94	0.042
ω_0	0.003	0.029	0.94	0.172
ω_1	0.000	0.005	0.95	0.073
ω_2	0.002	0.003	0.95	0.054
κ_0	-0.004	0.019	0.95	0.138
κ_1	0.001	0.001	0.95	0.041
κ_2	0.002	0.001	0.95	0.040
σ	-0.001	0.000	0.95	0.010

Corresponding GOF test: Type I error = 0.039.

4.3 Power of the Goodness-of-fit Test

In addition to the type I error, we also assessed the power of the GOF to detect the presence of model mis-specification for various departures from the assumed likelihood model. The results, presented in Table 3, show that the goodness-of-fit test has moderate to high power to detect certain forms of model misspecification for both binary and continuous exposures.

Table 3: Goodness-of-fit Test: Power

Misspecified Model	Missing covariates [a]	Parameter Values [a]	Power
	(1) *Binary Exposure*		
$q(X, Z, L)$	X^2	1.5	0.15
	Z, $X \times Z$	-0.6, 1.5	0.41
$t(L)$	L_2^2	1.5	0.40
	$L_1 \times L_2$	0.7	0.03
	L_2	1.5	0.89
	(2) *Continuous Exposure*		
$q(X, Z, L)$	X^2	-0.4	0.95
	Z, $X \times Z$	0.6, -1.5	0.62
$t(L)$	L_2^2	0.6	0.43
	$L_1 \times L_2$	0.6	0.06
	L_2	0.6	0.14

[a] Covariates (with corresponding parameter values) used in the generated model, but omitted in the fitted model.

5 Conclusion

In this paper, we presented a new parametrization for a logistic SNMM for a binary outcome and we proposed a corresponding maximum likelihood approach for estimation. Our approach builds upon the theoretical frameworks of VG and RR. Unlike VG, and similar to RR, our approach is guaranteed to always be congenial. However, unlike RR, we obviate the need to numerically solve an integral equation, which can be computationally cumbersome and is not easily scalable with the dimension of the exposure X. In addition, a key attraction of our approach is that it is readily implemented using standard statistical software. Our simulation results confirm the good performance of the proposed approach.

We also proposed a GOF test for the likelihood model, which is normal with mean zero only if the likelihood is correctly specified. The GOF statistic is based on an influence function for γ in a model where the likelihood is otherwise unrestricted, and therefore, the statistic naturally accounts for variability of all unknown parameters under the null of no model misspecification.

Our simulations showed that the proposed GOF is quite conservative in the settings we considered. Furthermore, the power of the test statistic to detect certain departures from the assumed model was either moderate or high, except for two exceptions. The poor performance of the GOF for these two cases may be a reflection of the conservative variance used to standardize the statistic. The main advantage of the current GOF is in the simplicity of the proposed standardization, however it appears to sometimes be overly conservative. In future work, we plan to further study the performance of the GOF statistic when standardized by a consistent estimator of its variance, which was not considered in the foregoing.

The method described here assumes random sample, as a straightforward extension will be to use inverse-probability weighting (IPW) using weihgts of selecting into a case-control study, which is well-known by design. However, this approach may be potentially efficient. More efficient methods for case-control studies similar to the one we proposed here will be discussed elsewhere.

Acknowledgments: We thank the reviewers for their resourceful comments and valuable suggestions, which have improved both the content and the presentation of this paper.
R. A. Matsouaka was supported by NIH grants 1R01MD006064 and 1R21HD066312 (Theresa Osypuk, PI) and E. Tchetgen Tchetgen was supported by NIH grants 1R21ES019712-01 and R01AI104459-01A1.

References

Didelez, V., Meng, S. and Sheehan, N. A. (2010). "Assumptions of IV methods for observational epidemiology", *Statist. Sci.*, **25**: 22-40

Meng, X. L. (1994), "Multiple-imputation inferences with uncongenial sources of input (with discussion)," *Statist. Sci.*, **9**: 538-573

Neyman, J (1923), "Sur les applications de la thóries des probabilités aux expériences agricoles: essai des principes," Excerpts reprinted in English (D. Dabrowska and T. Speed, Trans.) in *Statist. Sci.*, **5**: 463-472

Richardson, T. S., Evans, R. J., and Robins, J. M. (2011), "Transparent parametrizations of potential outcome models (with discussion)," in Bayesian Statistics 9, J.M. Bernardo, M.J. Bayarri, J. O. Berger, A. P. Dawid, D. Hekerman, A. F. M. Smith and M. West (Eds.), pp. 569-610.

Richardson, T. S. and Robins, J. M. (2010), "Analysis of the binary instrumental variable model," Heuristics, Probability and Causality: a tribute to Judea Pearl. R. Dechter, H. Geffner, and J. Halpern, Editors, College Publications, UK, pp. 415-444 (http://bayes.cs.ucla.edu/TRIBUTE/festschrift-complete.pdf)

Robins, J. M. (1986), "A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect," *Math. Mod.*, **7**: 1393-.

Robins, J. M. (1989), "The Analysis of Randomised and Non-Randomised AIDS Treatment Trials Using a New Approach To Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, eds. L. Sechrest, H. Freeman, and A. Mulley, Washington, DC: US Public Health Service, National Center for Health Services Research, pp. 113-159.

Robins, J. M. (1994), "The Analysis of Randomised and Non-Randomised AIDS Treatment Trials Using a New Approach To Causal Inference in Longitudinal Studies," *Comm. Statist. - Theory and Methods*, **23**: 2379-2412.

Robins, J. M., and 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.

Rubin, D. B. (1978), "Bayesian inference for causal effects: the role of randomization" *Ann. Statist.*, **6**, 34-58.

Tchetgen Tchetgen E., Robins J.M., Rotnitsky A. (2010). "On doubly robust estimation in a semiparametric odds ratio model" *Biometrika*, **97**(1):171-180.

Tsiatis, A. A. "Semiparametric Theory and Missing Data", Springer, New York

Vansteelandt, S. and Goetghebeur, E. (2003), "Causal Inference with Generalized Structural Mean Models", *J. R. Stat. Soc.*, Series B, **65**, 817-835.

Vansteelandt, S., Bowden, J, Babanezhad, M., and Goetghebeur, E. (2011), "On instrumental variable estimation of causal odds ratios ", *Statist. Sci.*, **26**, 403-422.