Received: 8 June 2022 | Accepted: 5 July 2022

DOI: 10.1111/biom.13790

DISCUSSION

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

# Discussion on "Instrumental variable estimation of the causal hazard ratio," by Linbo Wang, Eric Tchetgen Tchetgen, Torben Martinussen, and Stijn Vansteelandt

**Benjamin R. Baer** | **Robert L. Strawderman** | **Ashkan Ertefaie**

Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, USA

**Correspondence**
Ashkan Ertefaie, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, USA.
Email: ashkan_ertefaie@urmc.rochester.edu

**Abstract**
We propose and study an augmented variant of the estimator proposed by Wang, Tchetgen Tchetgen, Martinussen, and Vansteelandt.

**KEYWORDS**
causal inference, instrumental variable, unmeasured confounders

## 1 | INTRODUCTION

We thank the editor for the opportunity to discuss the article by Wang, Tchetgen Tchetgen, Martinussen, and Vansteelandt (henceforth WTTMV). We congratulate WTTMV on their interesting work, who argue that the wide use of the Cox proportional hazard model warrants study of the causal hazard ratio. We agree, noting here that the causal hazard ratio—defined as the ratio of two marginal hazards, one for each counterfactual survival time—can be difficult to interpret (Hernán, 2010; Martinussen et al., 2020).

WTTMV propose a plug-in estimator based on an identification result motivated by a clever simplification of a weighted Cox influence function. In this commentary, we consider improved estimation relative to this estimator. We freely use the WTTMV notation and throughout adopt all of their assumptions and regularity conditions.

## 2 | AN ALTERNATIVE ESTIMATOR

Theorem 2 in Wang et al. (2022) shows that the causal hazard ratio equals

$$\psi = \log \frac{\mathbb{E} \int dN(y)(-D)\omega_0(Z,X)\left\{m(1)\gamma_1(y) - \gamma_2^m(y)\right\}}{\mathbb{E} \int dN(y)(1-D)\omega_0(Z,X)\left\{m(0)\gamma_1(y) - \gamma_2^m(y)\right\}}$$
$$= \log \frac{\tau_1}{\tau_0} \tag{1}$$

under their assumptions. For simplicity, let $\tau_1$ denote the numerator and $\tau_0$ denote the denominator. We aim to find an improved estimator by making use of an efficient influence function (Hines et al., 2022). The WTTMV Cox marginal structural model is semiparametric, so the space of full data influence functions is not a singleton. This complicates the calculation of the observed data efficient influence function (Scheike et al., 2022; Tsiatis, 2006). Instead, we find the efficient influence function in an unrestricted and hence nonparametric model (Jiang et al., 2022) then study the properties of an associated estimator in the WTTMV semiparametric model.

### 2.1 | The nonparametric efficient influence function

The efficient influence function for estimating the causal hazard ratio in a nonparametric model equals

 

$$\frac{\mathrm{EIF}(\tau_1)\tau_0 - \tau_1 \mathrm{EIF}(\tau_0)}{\tau_0 \tau_1}, \tag{2}$$

where EIF represents the efficient influence function in a nonparametric model of its argument, by the quotient rule for canonical gradients (Hines et al., 2022; Kennedy, 2022). Thus, the problem of finding the nonparametric efficient influence function for estimating the estimand reduces to that of finding the nonparametric efficient influence function for estimating $\tau_1$ and $\tau_0$. For convenience, we only consider the numerator $\tau_1$, as the calculations for $\tau_0$ are analogous.

**Proposition 1.** *The efficient influence function $D^*(\mathcal{O})$ for estimating $\tau_1$ in a nonparametric model equals*

$$\begin{aligned}
&-\int dN(y)D\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\zeta_1(y)\\
&-\mathbb{E}\left\{dN(y)D\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\zeta_1(y)\right\}\\
&-\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\zeta_1(y)\mathbb{E}\{dN(y)D\mid X,Z\}\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{dN(y)D\mid X,Z=z\}\zeta_1(y)\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{dN(y)D\mid X,Z=z\}\zeta_1(y)\\
&\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\{D-\mathbb{E}(D\mid Z,X)\}\\
&+\mathbb{E}\left\{dN(y)D\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\right\}m(1)\Phi_{d4}(y;\mathcal{O})\\
&-\mathbb{E}\left\{dN(y)D\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\right\}\Phi_{d5}(y;\mathcal{O}),
\end{aligned} \tag{3}$$

*where*

$$\begin{aligned}
\Phi_{d5}(y;\mathcal{O}) = \Bigg[&m(D)I(Y\ge y)\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\\
&-\mathbb{E}\left\{m(D)I(Y\ge y)\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\right\}\Bigg]\\
&-\Bigg[\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\mathbb{E}\{m(D)I(Y\ge y)\mid X,Z\}\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{m(D)I(Y\ge y)\mid X,Z=z\}\Bigg]\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{m(D)I(Y\ge y)\mid X,Z=z\}\\
&\frac{2Z-1}{\delta^D(X)f(Z\mid X)}\{D-\mathbb{E}(D\mid Z,X)\},
\end{aligned} \tag{4}$$

$$\begin{aligned}
\Phi_{d4}(y;\mathcal{O}) = \Bigg[&I(Y\ge y)\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\\
&-\mathbb{E}\left\{I(Y\ge y)\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\right\}\Bigg]\\
&-\Bigg[\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\mathbb{E}\{I(Y\ge y)\mid X,Z\}\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{I(Y\ge y)\mid X,Z=z\}\Bigg]\\
&-\sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{I(Y\ge y)\mid X,Z=z\}\\
&\frac{2Z-1}{f(Z\mid X)\delta^D(X)}\{D-\mathbb{E}(D\mid Z,X)\},
\end{aligned} \tag{5}$$

*and $\zeta_1(y) = m(1)\gamma_1(y) - \gamma_2^m(y)$ and $\mathcal{O}$ denotes the observed data.*

The proof of this result is provided in Web Appendix A in the Supporting Information.

We remark here that under the nonparametric model, we are not assuming that the ratio of marginal hazards follows WTTMV's Cox proportional hazards model. As such, the resulting ratio $\tau_1/\tau_0$ under the nonparametric model does not reduce to $e^{\psi}$ but rather to $E_1(S_0^Y(Y))/E_0(S_1^Y(Y))$, where $E_a(g(Y))$ is calculated assuming that $Y$ follows a distribution governed by the density $f_a(y) = \lambda_a^Y(y)S_a^Y(y)$. Under WTTMV's model assumption, this expression reduces to $e^{\psi}$ as required.

## 2.2 | The semiparametric estimator

In this section, we define a simplified estimating equation motivated by the nonparametric efficient influence function for estimating $\psi = \log \tau_1/\tau_0$. Although the influence function in the previous section is efficient in a nonparametric model, it is still a (not necessarily efficient) influence function in WTTMV's model (Tsiatis, 2006, Chapter 4).

The influence functions in the previous section for the terms $\tau_1$ and $\tau_0$ have the form $s_1(\mathcal{O}) - \tau_1$ and $s_0(\mathcal{O}) - \tau_0$, where $\mathcal{O}$ denotes the observed data and $s_1, s_0$ are functions. Thus, the stated influence function of $\log \tau_1/\tau_0$ may be used to construct the estimating equation

$$0 = \mathbb{P}_n \frac{(s_1(\mathcal{O}) - \hat{\tau}_1)\hat{\tau}_0 - (s_2(\mathcal{O}) - \hat{\tau}_0)\hat{\tau}_1}{\hat{\tau}_0\hat{\tau}_1}, \tag{6}$$

$$\leftrightarrow \quad \log \hat{\tau}_1/\hat{\tau}_0 = \log \mathbb{P}_n s_1(\mathcal{O})/\mathbb{P}_n s_0(\mathcal{O}). \tag{7}$$

The estimator for $\psi$ conveniently decouples the estimation of $\tau_1$ and $\tau_0$.

The last two terms in the nonparametric efficient influence function for estimating $\tau_1$, in Proposition 1, lead to efficiency gains due to projection onto the nuisance tangent spaces for $\gamma_1$ and $\gamma_2^m$. As a simplification, we consider the estimating equation defined by removing these terms and hence only achieving efficiency gains due to the weight $(2Z-1)/\{f(Z \mid X)\delta^D(X)\}$. We additionally simplify the estimating equation by evaluating some terms at $Z=0$, following Cui and Tchetgen Tchetgen (2021). That is, we consider the estimator

$$\hat{\tau}_1 = -\mathbb{P}_n\Big[\delta^D(X)D\hat{\omega}_0(X,Z)\hat{\zeta}_1(Y) - \hat{\omega}_0(X,Z)\hat{\alpha}_1(X,0) + \hat{\beta}_1(X) - \hat{\omega}_0(X,Z)\hat{\beta}_1(X)\big\{D-\hat{E}(X,0)\big\}\Big], \quad (8)$$

for $\omega_0(X,Z) = \frac{2Z-1}{f(Z|X)\delta^D(X)}$, $\zeta_1(y) = m(1)\gamma_1(y) - \gamma_2^m(y)$, $\alpha_1(X,Z) = \mathbb{E}\{\Delta D\zeta_1(Y) \mid X, Z\}$, $\beta_1(X) = \sum_z \frac{2z-1}{\delta^D(X)}\mathbb{E}\{\Delta D\zeta_1(Y) \mid X, Z=z\}$, and $E(X,Z) = \widehat{\mathbb{E}}\{D \mid X, Z\}$. We likewise define an estimator $\hat{\tau}_0$ for $\tau_0$, with $\zeta_0, \alpha_0$, and $\beta_0$ defined analogously. We describe these estimators as the *augmented* estimators.

In the next section, we elaborate on the estimation of the nuisance parameters. After, we show that the augmented estimators are multiply robust.

## 2.3 | Nuisance parameter estimation

The instrumental variable propensity score $f$ and the expectation $E$ are estimated using parametric maximum likelihood. The estimated difference $\hat{\delta}^D$ solves the estimating equation

$$\frac{2Z-1}{\hat{f}(Z \mid X)}X\big\{D - Z\delta^D(X) - \hat{E}(X,0)\big\}, \quad (9)$$

following Appendix H in Cui and Tchetgen Tchetgen (2021). This estimator is doubly robust, as stated below.

**Lemma 1.** *The estimator $\hat{\delta}^D(X)$ is consistent if either (1) $f$ is correctly specified or (2) $E$ is correctly specified.*

This follows from Appendix H in Cui and Tchetgen Tchetgen (2021). The estimated weight $\hat{\omega}_0(X,Z) = (2Z-1)/\{\hat{f}(Z \mid X)\hat{\delta}^D(X)\}$ is a plug-in estimator. The estimator $\hat{\gamma}_1(y)$ solves the estimating equation

$$\hat{\omega}_0(X,Z)I(Y \geq y) - \hat{\omega}_0(X,Z)\hat{\alpha}_\gamma(y;X,0) + \hat{\beta}_\gamma(y;X) - \hat{\omega}_0(X,Z)\hat{\beta}_\gamma(y;X)\big\{D - \hat{E}(X,0)\big\} \quad (10)$$

for each $y$, where $\alpha_\gamma(y;X,Z) = \mathbb{E}[I(Y \geq y) \mid X, Z]$ and $\beta_\gamma(y;X) = \mathbb{E}[\hat{\omega}_0(X,Z)I(Y \geq y) \mid X]$. This estimator is based on the efficient influence function for estimating $\gamma_1(y)$ and is multiply robust, as stated below.

**Lemma 2.** *The estimator $\hat{\gamma}_1(y)$ is consistent if either (1) $f$ and $\delta^D$ are correctly specified, (2) $f$ and $\beta_\gamma$ are correctly specified, or (3) $\delta^D$, $\beta_\gamma$, $\alpha_\gamma$, and $E$ are correctly specified.*

This follows from a variant of Theorem H.1 in Cui and Tchetgen Tchetgen (2021).

We estimate $\alpha_\gamma(y;X,Z)$ using parametric maximum likelihood for each $y$. We estimate $\beta_\gamma(y;X)$ as the solution to the estimating equation

$$\frac{2Z-1}{\hat{f}(Z \mid X)}X\big[I(Y \geq y) - \hat{\alpha}_\gamma(y;X,0) - \beta_\gamma(y;X)\{D - \hat{E}(X,0)\}\big] \quad (11)$$

for each $y$. This estimator is also multiply robust, stated below.

**Lemma 3.** *The estimator $\hat{\beta}_\gamma(y;X)$ is consistent if either (1) $f$ is correctly specified or (2) $\alpha_\gamma$ and $E$ are correctly specified.*

This follows from Appendix H in Cui and Tchetgen Tchetgen (2021).

We similarly estimate $\gamma_2^m(y)$ using a multiply robust estimator, along with its nuisance parameters ($\alpha_\gamma^m$ and $\beta_\gamma^m$). The terms $\alpha_d(X,Z)$ are estimated using parametric maximum likelihood, for each $d=0,1$. The terms $\beta_d(X)$ are estimated in the same manner as $\beta_\gamma(y;X)$, for each $d=0,1$.

## 2.4 | Multiple robustness

In this section, we study the robustness of $\log(\hat{\tau}_1/\hat{\tau}_0)$ to misspecification of the parametric form of the nuisance parameters. The estimator $\log(\hat{\tau}_1/\hat{\tau}_0)$ is consistent and asymptotically normal if $(\hat{\tau}_1, \hat{\tau}_0)$ is consistent and asymptotic normal, as long as $\tau_0 \neq 0$.

We consider data generated according to three models, stated below.

$\mathcal{M}_1$ : Models for $f(Z \mid X), \delta^D(X)$ are correctly specified;
$\mathcal{M}_2$ : Models for $f(Z \mid X)$, $\beta_d(X)$ $(d=0,1)$, $\beta_\gamma(y;X)$, and $\beta_\gamma^m(y;X)$ are correctly specified;

**TABLE 1** The bias and (Monte Carlo) standard error of the plug-in and augmented estimators when $f$ is correctly and incorrectly specified

| | Bias | Standard error |
|---|---|---|
| Plug-in (correct spec.) | 0.002 | 0.23 |
| Plug-in (incorrect spec.) | 0.176 | 0.13 |
| Plug-in (oracle) | 0.002 | 0.22 |
| Augmented (correct spec.) | −0.002 | 0.26 |
| Augmented (incorrect spec.) | 0.004 | 0.18 |

$\mathcal{M}_3$ : Models for $\delta^D(X)$, $\beta_d(X)$ $(d = 0, 1)$, $\beta_\gamma(y; X)$, $\beta_\gamma^m(y; X)$, $\alpha_d(X, 0)$ $(d = 0, 1)$, $\alpha_\gamma(y; X, 0)$, $\alpha_\gamma^m(y; X, 0)$, and $E(X, 0)$ are correctly specified.

The following proposition states that the bivariate estimator $(\hat{\tau}_0, \hat{\tau}_1)$ is consistent and asymptotically normal if any of these three models hold.

**Proposition 2.** *Assume that the nuisance parameters (listed above) are parametrically specified. The estimator $(\hat{\tau}_0, \hat{\tau}_1)$ is a consistent and asymptotically normally estimator of $(\tau_0, \tau_1)$ under the model $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$.*

The proof is given in Web Appendix B in the Supporting Information.

Asymptotic normality of the estimator for $\psi$ then follows from the delta method.

## 3 | SIMULATION STUDY

In this section, we consider the finite sample performance of the proposed multiply robust augmented estimator compared to the plug-in estimator in Wang et al. (2022). We sampled from nearly the same distribution as WTTMV in their first setup. We drew $X_2$ and $X_3$ as i.i.d. Bernoulli$(1, 1/2)$ and correspondingly updated the distribution function of $T \mid X, U$ to use the Laplace transform of a Bernoulli rather than Exponential distribution. We also updated the coefficients of $f$ from $(-1/\lambda_2, 1, 0)$ to $(-1/\lambda_2, 3, -2)$ to increase the dependency on $X_2, X_3$. The use of discrete covariates ensures correct specification of each nuisance parameter.

Imposing Assumption $A5$ in WTTMV, we set $\lambda = 1$ and $n = 1000$. In Table 1, we compare the plug-in estimator to the augmented estimator. (Note that we updated the plug-in estimator to use the doubly robust (and consistent here) estimator of $\delta^D$ described in Section 2.3.) We considered both $f$ to be correctly specified, using $X_2$ and $X_3$ as covariates, and $f$ to be incorrectly specified, using an intercept only. Additionally, we considered the oracle plug-in estimator leveraging the true $f$ and $\delta^D$.

With correctly specified terms so that model $\mathcal{M}_1$ holds, the bias and standard error of the plug-in and augmented estimators were comparable. The low bias agrees with the theory that shows that both the plug-in and augmented estimators are consistent.

With $f$ incorrectly specified, model $\mathcal{M}_3$ holds but model $\mathcal{M}_1$ does not. Since the theory for the plug-in estimator relies on $f$ and $\delta^D$ being consistent, the plug-in estimator has a high bias. However, the multiple robustness property of the augmented estimator allows it to maintain low bias despite $f$ being incorrectly specified.

Interestingly, the augmented estimator under misspecification outperforms the augmented estimator under correct specification. A possible explanation is that variability in the estimation of $f$ noticeably propagates to variability in the augmented estimator due to $f$ being nested within each nuisance parameter.

Code to reproduce the simulation is available in the Supporting Information.

## 4 | CONCLUSION

The intersection of causal inference and survival analysis is at the core of many pressing scientific problems. WTTMV leveraged modern instrumental variable identification strategies (Wang & Tchetgen Tchetgen, 2018) to identify the hazard ratio in a Cox marginal structural model. To deliver improved estimation of the plug-in estimator, we made significant use of calculations provided in later work by a subset of these authors (Cui & Tchetgen Tchetgen, 2021).

### ORCID

*Benjamin R. Baer* https://orcid.org/0000-0002-9662-9377
*Robert L. Strawderman* https://orcid.org/0000-0002-6624-0272
*Ashkan Ertefaie* https://orcid.org/0000-0003-2611-9512

### REFERENCES

Cui, Y. & Tchetgen Tchetgen, E. (2021) A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association*, 116(533), 162–173.

Hernán, M.A. (2010) The hazards of hazard ratios. *Epidemiology*, 21(1), 13.

Hines, O., Dukes, O., Diaz-Ordaz, K. & Vansteelandt, S. (2022) Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 1–13, in press.

Jiang, Z., Yang, S. & Ding, P. (2022) Multiply robust estimation of causal effects under principal ignorability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1–23, in press.

Kennedy, E.H. (2022) Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 1–37.

Martinussen, T., Vansteelandt, S. & Andersen, P.K. (2020) Subtleties in the interpretation of hazard contrasts. *Lifetime Data Analysis*, 26(4), 833–855.

Scheike, T.H., Martinussen, T. & Ozenne, B. (2022) Efficient estimation in the Fine and Gray model. *Journal of the American Statistical Association*, 1–23, in press.

Tsiatis, A.A. (2006) *Semiparametric Theory and Missing Data*. New York: Springer.

Wang, L. & Tchetgen Tchetgen, E. (2018) Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3), 531–550.

Wang, L., Tchetgen Tchetgen, E., Martinussen, T. & Vansteelandt, S. (2022) Instrumental variable estimation of the causal hazard ratio. *Biometrics*, 1–26, in press.

## SUPPORTING INFORMATION

Web Appendix A, referenced in Section 2.1, Web Appendix B referenced in Section 2.4, and R programs implementing the proposed method are available at the Biometrics website on Wiley Online Library.

**How to cite this article:** Baer, B.R., Strawderman, R.L. & Ertefaie, A. (2023) Discussion on "Instrumental variable estimation of the causal hazard ratio," by Linbo Wang, Eric Tchetgen Tchetgen, Torben Martinussen, and Stijn Vansteelandt. *Biometrics*, 79, 554–558. https://doi.org/10.1111/biom.13790