12e édition
Rendez-vous de la recherche pharmaceutique
En présence le jeudi 7 décembre 2023

TNDDR : Efficient and doubly robust estimation of vaccine effectiveness under the test-negative design

Cong Jiang [1] Mireille Schnitzer [1] Denis Talbot [2]

[1]Université de Montréal [2]Université Laval

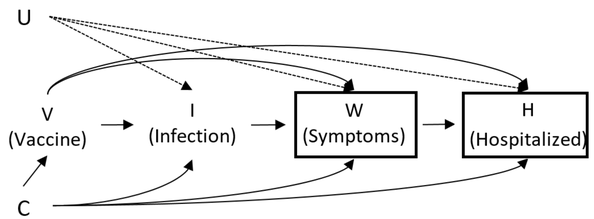
Faculté de pharmacie
Université de Montréal et du monde.

Context : Coronavirus disease 2019 (COVID-19) pandemic was caused by SARS-CoV-2 infection. Scientific questions regarding the effectiveness of different policy and personal measures abounded. Answers were and are needed on short time frames.

- Bulk processing of polymerase chain reaction (PCR) tests made mass testing possible.
- Many countries increased testing in early 2020. In Quebec, community PCR testing halted in January 2022, replaced by antigen tests with optional result reporting.
- Testing was largely driven by symptomatic status, contact tracing, high risk assessment, or administrative reasons.

The study design and data structure

The test-negative design (TND) is an observational study design that compares outcomes in people *who had symptoms characteristic of an infectious disease of interest and obtained a test* to estimate the effectiveness of vaccination meant to protect against disease outcomes caused by the infection.



Complete data from the underlying population are :

$$\boldsymbol{Z}^{\boldsymbol{C}} = (\boldsymbol{C}, V, I, W, H)$$

with probability distribution $\mathbb{P}$ under simple random sampling (SRS).

Define $S = \mathbb{I}(I \neq 0, W = 1, H = 1)$, the presence of inclusion criteria for the test-negative design.

The observed data are :

$$\boldsymbol{Z} = (\boldsymbol{C}, V, I)\mathbb{I}(S = 1)$$

with probability distribution $\mathbb{P}(\boldsymbol{Z} \mid S = 1)$ or $\mathbb{P}_{\text{TND}}(\boldsymbol{Z})$. The relevant outcome is *the hospitalized symptomatic SARS- CoV-2 infection* $Y = HW\mathbb{I}(I = 2)$.

Target estimand of vaccine effectiveness

Estimate a marginal risk ratio (mRR) under $\mathbb{P}$,

$$\psi_{mRR} = \frac{\mathbb{E}\{\mathbb{P}(Y = 1 \mid \boldsymbol{C}, V = v)\}}{\mathbb{E}\{\mathbb{P}(Y = 1 \mid \boldsymbol{C}, V = v_0)\}},$$

and the Vaccine Effectiveness (VE) is $1 - \psi_{mRR}$.

In past work, (Schnitzer 2022) proved the identifiability of this parameter and proposed an inverse probability of vaccination weighting (IPW) procedure that can estimate it.

Contributions

In this work, we derived

1. Efficient influence function and von Mises expansion of the ψ_{mRR}.
2. A one-step doubly robust and locally efficient estimator that can be implemented with machine learning (ML) methods for the nuisance functions converging at $o_{\mathbb{P}_{\text{TND}}}(n^{-1/4})$ rates.
3. TNDDR (TND doubly robust), which utilizes sample splitting and incorporate ML to estimate VE.

Efficient influence function

Define a key statistical estimand as

$$\psi_v := \psi_v(\mathbb{P}_{\text{TND}}) = \mathbb{E}_{\text{TND}}\left[\mu_v(\boldsymbol{c})\omega_v(\boldsymbol{c})\right], \tag{1}$$

with the specific form of $\omega(\boldsymbol{c}) = [1 - m(\boldsymbol{c})]/[1 - \mu_v(\boldsymbol{c})]$, where $\mu_v(\boldsymbol{c}) := \mathbb{P}_{\text{TND}}(Y = 1 \mid V = v, \boldsymbol{C} = \boldsymbol{c})$ and $m(\boldsymbol{c}) := \mathbb{P}_{\text{TND}}(Y = 1 \mid \boldsymbol{C} = \boldsymbol{c})$, and define $\pi_v^0(\boldsymbol{c}) := \mathbb{P}_{\text{TND}}(V = v \mid \boldsymbol{C} = \boldsymbol{c}, Y = 0, S = 1)$. The efficient influence function of ψ_v is

$\varphi_v(\boldsymbol{Z}, \mathbb{P}_{\text{TND}}) =$

$$\frac{\mathbb{I}(Y = 1, V = v)}{\pi_v^0(\boldsymbol{C})} - \mu_v(\boldsymbol{C})\left\{\frac{\mathbb{I}\left(Y = 0, S = 1\right)\left[\mathbb{I}(V = v) - \pi_v^0(\boldsymbol{C})\right]}{\pi_v^0(\boldsymbol{C})[1 - \mu_v(\boldsymbol{C})]}\right\} - \psi_v(\mathbb{P}_{\text{TND}})$$

$odds_{\text{TND}}(Y \mid V = v, \boldsymbol{C}) := \mu_v(\boldsymbol{C})/[1 - \mu_v(\boldsymbol{C})]$ is the odds under $\mathbb{P}_{\text{TND}}$ of $Y = 1$ for $V = v$ given $\boldsymbol{C}$.

Theorem 1

Define an estimator for ψ_v as $\hat{\psi}_v := \mathbb{P}_{\text{TND},n}[\phi_v(\boldsymbol{Z}; \hat{\pi}_v^0, \hat{\mu}_v)]$, where $\phi_v(\boldsymbol{Z}; \hat{\pi}_v^0, \hat{\mu}_v) =$

$$\frac{\mathbb{I}(Y = 1, V = v)}{\hat{\pi}_v^0(\boldsymbol{c})} - \hat{\mu}_v(\boldsymbol{c})\frac{\mathbb{I}\left(Y = 0, S = 1\right)\left[\mathbb{I}(V = v) - \hat{\pi}_v^0(\boldsymbol{c})\right]}{\hat{\pi}_v^0(\boldsymbol{c})[1 - \hat{\mu}_v(\boldsymbol{c})]}.$$

Under the identification assumptions and some conditions, then the estimator $\hat{\psi}_v := \mathbb{P}_{\text{TND},n}[\phi_v(\boldsymbol{Z}; \hat{\pi}_v^0, \hat{\mu}_v)]$ satisfies

$$\hat{\psi}_v - \psi_v = \left|R_2^v(\hat{\mathbb{P}}_{\text{TND}}, \mathbb{P}_{\text{TND}})\right| + (\mathbb{P}_{\text{TND},n} - \mathbb{P}_{\text{TND}})[\varphi_v(\boldsymbol{Z}; \mathbb{P}_{\text{TND}})] + o_{\mathbb{P}_{\text{TND}}}\left(1/\sqrt{n}\right),$$

Terms of decomposition :
$\left|R_2^v(\hat{\mathbb{P}}_{\text{TND}}, \mathbb{P}_{\text{TND}})\right|$ 2nd-order bias term, bounded by the sum of two products ;
$(\mathbb{P}_{\text{TND},n} - \mathbb{P}_{\text{TND}})[\varphi_v(\boldsymbol{Z}; \mathbb{P}_{\text{TND}})]$ is a sample average of a fixed function → CLT

If the 2nd-order remainder term is negligible, i.e., of order $o_{\mathbb{P}_{\text{TND}}}(1/\sqrt{n})$, then

$$\sqrt{n}(\hat{\psi}_v - \psi_v) = \sqrt{n}(\mathbb{P}_{\text{TND},n} - \mathbb{P}_{\text{TND}})[\varphi_v(\boldsymbol{Z}; \mathbb{P}_{\text{TND}})] + o_{\mathbb{P}_{\text{TND}}}(1), \tag{2}$$

that is, $\hat{\psi}_v$ is $\sqrt{n}$-consistent and asymptotically normal.

Q : Under what conditions does the 2nd-order term become negligible ?

$\sqrt{n}$-consistent and asymptotically normal & double robustness

Under the assumptions in Theorem 1 and the following two conditions :
1. $\left\|\hat{\pi}_v^0 - \pi_v^0\right\| = o_{\mathbb{P}_{\text{TND}}}\left(n^{-1/4}\right)$; 2. $\|\hat{\mu}_v - \mu_v\| = O_{\mathbb{P}_{\text{TND}}}\left(n^{-1/4}\right)$;
The proposed the estimator $\hat{\psi}_v$ **is $\sqrt{n}$-consistent and asymptotically normal**, and the limiting distribution is $\sqrt{n}\left(\hat{\psi}_v - \psi_v\right) \rightsquigarrow \mathcal{N}\left(0, \text{var}\left(\varphi_v(Z)\right)\right)$, where $\text{var}\left(\varphi_v(Z)\right) = \mathbb{E}\left(\varphi_v^2(Z)\right)$.

Can use ML methods to estimate the nuisance functions at $n^{-1/4}$ rates to obtain $n^{-1/2}$ rates for $\hat{\psi}_v$.

Under the assumptions in Theorem 1, then the proposed estimator $\hat{\psi}_v$ **is doubly robust.**

That is, either the propensity score (PS) model is correctly specified, i.e., $\left\|\hat{\pi}_v^0 - \pi_v^0\right\| = o_{\mathbb{P}_{\text{TND}}}(1)$ and/or the outcome models are correctly specified, i.e., $\|\hat{\mu}_v - \mu_v\| = o_{\mathbb{P}_{\text{TND}}}(1)$, then the remainder term

$$\left|R_2^v(\hat{\mathbb{P}}_{\text{TND}}, \mathbb{P}_{\text{TND}})\right| = o_{\mathbb{P}_{\text{TND}}}(1),$$

and thus $\hat{\psi}_v$ is consistent.

Correctly specifying the PS and/or Outcome models results in eliminating the 2nd-order remainder term.

Three steps of TNDDR (TND doubly robust)

- Step 1 : Randomly split the observed TND data into $J \geq 2$ disjoint, evenly sized folds.
- Step 2 : For each $j = 1, 2, ..., J$, using the data that excludes the j-th fold (data in $\boldsymbol{Z}_{Q_{(-j)}}$), we use ML methods to estimate the nuisance parameters $\pi_v^0(\cdot)$ and $\mu_v(\cdot)$.
- Step 3 : Construct the efficient VE estimator i.e., $1 - \hat{\psi}_1/\hat{\psi}_0$ with the nuisance components estimated in Step 2, where for $v \in \{0, 1\}$,

$$\hat{\psi}_v = \frac{1}{n}\sum_{j=1}^{J}\sum_{k \in Q_j}\left\{\frac{\mathbb{I}(y_k = 1, v_k = v)}{\hat{\pi}_v^0\left(\boldsymbol{c}_k; z_{Q_{(-j)}}\right)}\hat{\mu}_v\left(\boldsymbol{c}_k; z_{Q_{(-j)}}\right)\frac{\mathbb{I}\left(y_k = 0, s_k = 1\right)\left[\mathbb{I}(v_k = v) - \hat{\pi}_v^0\left(\boldsymbol{c}_k; z_{Q_{(-j)}}\right)\right]}{\hat{\pi}_v^0\left(\boldsymbol{c}_k; z_{Q_{(-j)}}\right)\left[1 - \hat{\mu}_v\left(\boldsymbol{c}_k; z_{Q_{(-j)}}\right)\right]}\right\}$$

Therefore, the marginal RR is efficiently and double-robustly estimated by $\hat{\psi}_{mRR}^{eif} = \hat{\psi}_1/\hat{\psi}_0$.

Results from the INSPQ data

An administrative dataset of older people (aged $\geq$ 60y) in Québec, Canada (Institut national de santé publique du Québec, INSPQ). This study covered the timeframes characterized by the Omicron BA.5 dominance period, from July 3 to November 5, 2022.
Vaccination : $V = 1$ booster mRNA vaccine within 6 months (with at least 7 days before testing), $V = 0$ final dose 6+ months prior to testing.
Covariates $\boldsymbol{c}$ include : 1. Age groups ; 2. Gender (Female/Male) ; 3. Multimorbidity (Yes/No) : minimum two health conditions elevating COVID-19 susceptibility ; 4. Epidemiological observation timeframe.

Methods & Results	GLM	ML (MARS)
IPW	0.655 (0.610, 0.703)	0.659 (0.614, 0.707)
OutR	0.655 (0.609, 0.719)	0.669 (0.619, 0.793)
TNDDR	0.650 (0.602, 0.703)	0.646 (0.599, 0.696)
Logistic regression	0.655 (0.630, 0.680)	

Figure 1. Estimated ψ_{mRR} and 95% confidence intervals were obtained using IPW, OutR, TNDDR, and logistic regression, employing GLM and machine learning (MARS) for nuisance function computation.

Comparing hospitalized symptomatic SARS-CoV-2 infection for booster mRNA vaccine within 6 months (with at least 7 days before testing) versus final dose administered 6+ months before testing, TNDDR VE estimate : GLM : 35.0% (29.7%, 39.8%); MARS : 35.4% (30.4%, 40.1%).

Acknowledgements

This research is supported by a Project Grant from the Canadian Institutes of Health Research (CIHR) and Centre de Recherches Mathematiques StatLab Postdoctoral Fellowship.

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada
CENTRE DE RECHERCHES MATHÉMATIQUES

The arXiv link.