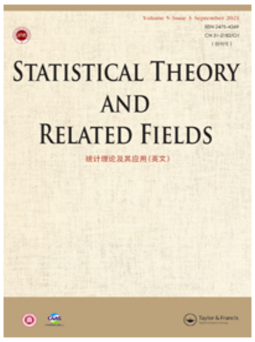

Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Comment: inference after covariate-adaptive randomisation: aspects of methodology and theory

Ting Ye & Yanyao Yi

To cite this article: Ting Ye & Yanyao Yi (2021) Comment: inference after covariate-adaptive randomisation: aspects of methodology and theory, Statistical Theory and Related Fields, 5:3, 194-195, DOI: 10.1080/24754269.2021.1905377

To link to this article: https://doi.org/10.1080/24754269.2021.1905377

Published online: 22 Mar 2021.

STATISTICAL THEORY AND RELATED FIELDS
2021, VOL. 5, NO. 3, 194–195
https://doi.org/10.1080/24754269.2021.1905377

Comment: inference after covariate-adaptive randomisation: aspects of methodology and theory

Ting Ye [a] and Yanyao Yi [b]

[a] Department of Statistics, University of Pennsylvania, Philadelphia, PA, USA; [b] Global Statistical Sciences, Eli Lilly and Company, Indianapolis, IN, USA

ARTICLE HISTORY Received 11 March 2021; Revised 14 March 2021; Accepted 16 March 2021

We first want to commend (Shao, 2021) for a timely paper that reviews the methodological and theoretical advances in statistical inference after covariate-adaptive randomisation in the last decade. The paper clearly presents the important considerations and pragmatic recommendations when analysing data obtained from covariate-adaptive randomisation, which provides principled guidelines for the practice.

The aim of our remaining comments is to extend the discussion on the invariance property in Shao (2021). That is, the asymptotic distribution of an estimator remains the same under different covariate-adaptive randomisation schemes. For ease of reading, we follow the notation in Shao (2021) whenever possible and focus on the case of two treatment arms (i.e., $k = 2$). The ideas can be extended to the case of multiple treatment arms.

For continuous or binary outcomes, Shao (2021) describes three post-stratified estimators for the population mean difference $\theta_0 = E(Y^{(2)} - Y^{(1)})$:

$$\hat\theta_S = \sum_{\mathbf{z}\in\mathcal{Z}} \frac{n(\mathbf{z})}{n}\{\overline{Y}_2(\mathbf{z}) - \overline{Y}_1(\mathbf{z})\},$$

$$\begin{aligned}\hat\theta_A = \sum_{\mathbf{z}\in\mathcal{Z}} \frac{n(\mathbf{z})}{n}\big[&\overline{Y}_2(\mathbf{z}) - \overline{Y}_1(\mathbf{z}) \\ &- \{\overline{\mathbf{U}}_2(\mathbf{z}) - \overline{\mathbf{U}}(\mathbf{z})\}\hat\beta_2(\mathbf{z}) \\ &+ \{\overline{\mathbf{U}}_1(\mathbf{z}) - \overline{\mathbf{U}}(\mathbf{z})\}\hat\beta_1(\mathbf{z})\big],\end{aligned}$$

$$\begin{aligned}\hat\theta_B = \sum_{\mathbf{z}\in\mathcal{Z}} \frac{n(\mathbf{z})}{n}\big[&\overline{Y}_2(\mathbf{z}) - \overline{Y}_1(\mathbf{z}) \\ &- \{\overline{\mathbf{U}}_2(\mathbf{z}) - \overline{\mathbf{U}}(\mathbf{z})\}\hat\beta(\mathbf{z}) + \{\overline{\mathbf{U}}_1(\mathbf{z}) - \overline{\mathbf{U}}(\mathbf{z})\}\hat\beta(\mathbf{z})\big],\end{aligned}$$

where $\mathcal{Z}$ is the support of $\mathbf{Z}_i$, and all other quantities are defined in Sections 5.2 and 6.1 of Shao (2021). These post-stratified estimators enjoy the invariance property; that is, their asymptotic distributions are not affected by the covariate-adaptive randomisation. This is a very appealing property as (i) the same inference procedure can be universally applied to different covariate-adaptive randomisation schemes; and (ii) valid inference of the treatment effect can be obtained when complicated covariate-adaptive randomisation schemes such as the Pocock and Simon's minimisation are employed.

It is well-known that the post-stratified estimator $\hat\theta_S$ is algebraically equivalent with the estimator of θ from fitting the following *working model*

$$\begin{aligned}&E(Y_i \mid A_i, \mathbf{Z}_i) \\ &= \alpha + \theta I(A_i = e_2) \\ &\quad + \sum_{t=1}^{2}\sum_{\mathbf{z}\in\mathcal{Z}_{-1}} I(A_i = e_t)\left\{I(\mathbf{Z}_i = \mathbf{z}) - \frac{n(\mathbf{z})}{n}\right\}\eta_t(\mathbf{z}),\end{aligned}$$

where $\alpha, \theta, \eta_1(\mathbf{z}), \eta_2(\mathbf{z})$ are unknown parameters, and $\mathcal{Z}_{-1}$ is the support of $\mathbf{Z}_i$ with one level dropped to avoid degeneracy. See, for example, Fuller (2009, Chapter 2.2.3). Analogously, $\hat\theta_A$ is algebraically equivalent with the estimator of θ from fitting the following *working model*

$$\begin{aligned}&E(Y_i \mid A_i, \mathbf{Z}_i, \mathbf{U}_i) \\ &= \alpha + \theta I(A_i = e_2) \\ &\quad + \sum_{t=1}^{2}\sum_{\mathbf{z}\in\mathcal{Z}_{-1}} I(A_i = e_t)\left\{I(\mathbf{Z}_i = \mathbf{z}) - \frac{n(\mathbf{z})}{n}\right\}\eta_t(\mathbf{z}) \\ &\quad + \sum_{t=1}^{2}\sum_{\mathbf{z}\in\mathcal{Z}} I(A_i = e_t) \\ &\quad \times \left\{I(\mathbf{Z}_i = \mathbf{z})\mathbf{U}_i - \frac{n(\mathbf{z})}{n}\overline{\mathbf{U}}(\mathbf{z})\right\}\beta_t(\mathbf{z}),\end{aligned}$$

and $\hat\theta_B$ is algebraically equivalent with the estimator of θ from fitting the following *working model*

$$\begin{aligned}&E(Y_i \mid A_i, \mathbf{Z}_i, \mathbf{U}_i) \\ &= \alpha + \theta I(A_i = e_2)\end{aligned}$$

CONTACT Ting Ye tingye@wharton.upenn.edu

$$+\sum_{t=1}^{2}\sum_{\mathbf{z}\in\mathcal{Z}_{-1}} I(A_i=e_t)\left\{I(\mathbf{Z}_i=\mathbf{z})-\frac{n(\mathbf{z})}{n}\right\}\eta_t(\mathbf{z})$$

$$+\sum_{\mathbf{z}\in\mathcal{Z}}\left\{I(\mathbf{Z}_i=\mathbf{z})\mathbf{U}_i-\frac{n(\mathbf{z})}{n}\overline{\mathbf{U}}(\mathbf{z})\right\}\beta(\mathbf{z}),$$

where $\beta(\mathbf{z}),\beta_1(\mathbf{z}),\beta_2(\mathbf{z})$ are unknown parameters.

These three invariant estimators $\hat\theta_S,\hat\theta_A,\hat\theta_B$ can each be obtained as the estimator of θ from fitting the following working model, with properly specified $\mathbf{W}_i$ and $\mathbf{V}_i$ being functions of $\mathbf{X}_i$, and $\overline{\mathbf{W}}$ and $\overline{\mathbf{V}}$ being their sample means,

$$\begin{aligned}&E(Y_i\mid A_i,\mathbf{W}_i,\mathbf{V}_i)\\&=\alpha+\theta I(A_i=e_2)\\&\quad+\sum_{t=1}^{2}I(A_i=e_t)(\mathbf{W}_i-\overline{\mathbf{W}})\lambda_{t\mathbf{W}}+(\mathbf{V}_i-\overline{\mathbf{V}})\lambda_{\mathbf{V}},\end{aligned}\tag{1}$$

where $\lambda_{1\mathbf{W}},\lambda_{2\mathbf{W}},\lambda_{\mathbf{V}}$ are unknown parameters. Here, the set of covariates included in $\mathbf{W}_i$ and $\mathbf{V}_i$ are non-overlapping to avoid degeneracy, where $\mathbf{W}_i$ has full interactions with A_i, while $\mathbf{V}_i$ does not have interactions with A_i. Specifically, $\hat\theta_S$ can be obtained with $\mathbf{W}_i=(I(\mathbf{Z}_i=\mathbf{z}),\mathbf{z}\in\mathcal{Z}_{-1})^{\mathrm{T}}$ being a column vector of all dummy variables for $\mathcal{Z}_{-1}$ and $\mathbf{V}_i$ being empty, $\hat\theta_A$ can be obtained with $\mathbf{W}_i=((I(\mathbf{Z}_i=\mathbf{z}),\mathbf{z}\in\mathcal{Z}_{-1}),(I(\mathbf{Z}_i=\mathbf{z})\mathbf{U}_i,\mathbf{z}\in\mathcal{Z}))^{\mathrm{T}}$ and $\mathbf{V}_i$ being empty, and $\hat\theta_B$ can be obtained with $\mathbf{W}_i=(I(\mathbf{Z}_i=\mathbf{z}),\mathbf{z}\in\mathcal{Z}_{-1})^{\mathrm{T}}$ and $\mathbf{V}_i=(I(\mathbf{Z}_i=\mathbf{z})\mathbf{U}_i,\mathbf{z}\in\mathcal{Z})^{\mathrm{T}}$.

In fact, model (1) defines a general class of estimators of the treatment effect θ_0,

$$\begin{aligned}\hat\theta&=\overline{Y}_2-\overline{Y}_1-(\overline{\mathbf{W}}_2-\overline{\mathbf{W}})\hat\lambda_{2\mathbf{W}}\\&\quad+(\overline{\mathbf{W}}_1-\overline{\mathbf{W}})\hat\lambda_{1\mathbf{W}}-(\overline{\mathbf{V}}_2-\overline{\mathbf{V}})\hat\lambda_{\mathbf{V}}+(\overline{\mathbf{V}}_1-\overline{\mathbf{V}})\hat\lambda_{\mathbf{V}},\end{aligned}\tag{2}$$

where $\hat\lambda_{1\mathbf{W}},\hat\lambda_{2\mathbf{W}},\hat\lambda_{\mathbf{V}}$ are the least squares estimates from fitting the working model (1). Similar to the proof of Theorem 2 in Ye et al. (2020), we can show that the class of estimators defined in (2) are consistent and asymptotically normal and have asymptotic distributions invariant to the covariate-adaptive randomisation schemes, as long as $\mathbf{W}_i$ includes the dummy variables for all joint levels of $\mathbf{Z}_i$ as a sub-vector. The key step in the proof is to show that for $t=1,2$,

$$\begin{aligned}&E\{Y_i^{(t)}-E(Y_i^{(t)})-(\mathbf{W}_i-E(\mathbf{W}_i))\lambda_{t\mathbf{W}0}\\&\quad-(\mathbf{V}_i-E(\mathbf{V}_i))\lambda_{\mathbf{V}0}\mid\mathbf{Z}_i\}=0,\end{aligned}\tag{3}$$

where $\lambda_{1\mathbf{W}0},\lambda_{2\mathbf{W}0}$, and $\lambda_{\mathbf{V}0}$ are the probability limits of $\hat\lambda_{1\mathbf{W}},\hat\lambda_{2\mathbf{W}},\hat\lambda_{\mathbf{V}}$ defined as

$$\begin{aligned}&(\alpha_0,\theta_0,\lambda_{\mathbf{V}0},\lambda_{1\mathbf{W}0},\lambda_{2\mathbf{W}0})\\&=\arg\min_{\alpha,\theta,\lambda_{\mathbf{V}},\lambda_{1\mathbf{W}},\lambda_{2\mathbf{W}}}E\left[\left\{Y_i-\alpha-\theta I(A_i=e_2)\right.\right.\\&\quad-\sum_{t=1}^{2}I(A_i=e_t)(\mathbf{W}_i-E(\mathbf{W}_i))\lambda_{t\mathbf{W}}\\&\quad\left.\left.-(\mathbf{V}_i-E(\mathbf{V}_i))\lambda_{\mathbf{V}}\right\}^2\right].\end{aligned}$$

Taking the derivatives and rearranging give $\alpha_0=E(Y^{(1)})$, $\theta_0=E(Y^{(2)})-E(Y^{(1)})$, and

$$\begin{aligned}&E\Big[\mathbf{W}_i\Big\{Y_i^{(t)}-E(Y_i^{(t)})-(\mathbf{W}_i-E(\mathbf{W}_i))\lambda_{t\mathbf{W}0}\\&\quad-(\mathbf{V}_i-E(\mathbf{V}_i))\lambda_{\mathbf{V}0}\Big\}\Big]=0.\end{aligned}$$

Because $\mathbf{Z}_i$ is discrete and $\mathbf{W}_i$ contains all joint levels of $\mathbf{Z}_i$ as a sub-vector, we have that (3) holds.

Equation (2) provides a more complete characterisation of the estimators satisfying the invariance property, compared with the ANHECOVA estimator in Ye et al. (2020) derived from (1) without the term $(\mathbf{V}_i-\overline{\mathbf{V}})\lambda_{\mathbf{V}}$. However, we should note that although the class of estimators defined in (2) enjoy the invariance property, they may be less efficient than the simple mean difference $\overline{Y}_2-\overline{Y}_1$ under some data generating process. This stresses the advantage of the ANHECOVA estimator in Ye et al. (2020) that is guaranteed to be more efficient than the simple mean difference $\overline{Y}_2-\overline{Y}_1$.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Ting Ye is a Postdoctoral Fellow at the Department of Statistics, The Wharton School, University of Pennsylvania. She obtained a Ph.D. in Statistics at the University of Wisconsin-Madison in 2019. Her research interest centers around developing pragmatic and robust statistical methods to analyze complex datasets, and advancing causal inference in biomedical and social sciences.

Yanyao Yi is a Research Scientist at the Global Statistical Sciences, Eli Lilly and Company. He obtained a Ph.D. in Statistics at the University of Wisconsin-Madison in 2019. His research mostly focuses on clinical trials.

ORCID

Ting Ye http://orcid.org/0000-0001-6009-641X

References

Fuller, W. A. (2009). *Sampling statistics*. Wiley.

Shao, J. (2021). Inference after covariate-adaptive randomisation: Aspects of methodology and theory. *Statistical Theory and Related Fields*, *5*. https://doi.org/10.1080/24754269.2021.1871873

Ye, T., Shao, J., Yi, Y., & Zhao, Q. (2020). Principles for covariate adjustment in analysing randomized clinical trials. arXiv:2009.11828.