The International Journal of Biostatistics

Volume 4, *Issue* 1 2008 *Article* 12

Rejoinder to Tan

Daniel B. Rubin, *University of California, Berkeley*
Mark J. van der Laan, *University of California, Berkeley*

Recommended Citation:
Rubin, Daniel B. and van der Laan, Mark J. (2008) "Rejoinder to Tan," *The International Journal of Biostatistics*: Vol. 4: Iss. 1, Article 12.
DOI: 10.2202/1557-4679.1116

Rejoinder to Tan

Daniel B. Rubin and Mark J. van der Laan

Abstract

We respond to several interesting points raised by Tan regarding our article.

We thank Tan for his comments on our covariate adjustment proposal, and for making several noteworthy contributions.

Tan initially points out that when the outcome regression model is linear, the estimators given in our paper can reduce to estimators previously presented by Tan himself and by Robins, Rotnitzky, and Zhao. We had not observed this, and it is an important clarification.

Tan also notes that one estimator we introduced has a desired double robustness property, meaning it will be accurate if either an outcome regression model or propensity score model is correctly specified. While superfluous for the randomized experiment setting of our paper, it is nice to know that our "improved local efficiency" can be achieved in observational studies, or experiments with noncompliance or loss to follow-up.

Additionally, Tan proposes a computationally convenient method for fitting the outcome regression model when the propensity score model is estimated. Again, although this is not strictly necessary for experiments, it is worthwhile because estimating the propensity scores can enhance efficiency even when they are known by design.

Tan finally notes that with binary outcomes, for which the outcome regression model $m_1(X;\alpha_1)$ wouldn't necessarily be linear in α_1, it can still be fruitful to reduce to the linear setting with the extension $m_1(X;\alpha_1,\beta_0,\beta_1) = \beta_0 + \beta_1 m_1(X;\alpha_1)$. A previous proposal had been to fit α_1 through standard methods with some $\hat{\alpha}_1$, and then fit (β_0,β_1) by treating $m_1(X;\hat{\alpha}_1)$ as a new covariate in an outcome regression model linear in (β_0,β_1). However, Tan observes that one could also view the extension as inducing a new outcome regression model, and that this nonlinear model could be fit jointly over the $(\alpha_1,\beta_0,\beta_1)$ parameter space. Ideally, the fit would be made to optimize performance for the parameter of interest μ_1, and a standard fit $\hat{\alpha}_1$ might perform poorly in this regard when the outcome regression model is misspecified.

So how should one then fit $(\alpha_1,\beta_0,\beta_1)$? It is for precisely these types of applications that we introduced our empirical efficiency maximization approach. When fitting an outcome regression, one must often confront many choices involving covariate transformations, variable/model selection, and (linear or nonlinear) models or extensions. Our technique provides a criterion for forming regression fits to minimize asymptotic variance for the resulting parameter estimator, and we hope this property is valuable for data analysis.