Re: Designing a Randomized Clinical Trial to Evaluate Personalized Medicine: A New Approach Based on Risk Prediction

Baker and Sargent (1) proposed using data from patients in the control arm of a randomized clinical trial (RCT) to develop a prognostic score using patient outcome and covariate information and then to compare the new treatment to control arms in subgroups of patients defined by ranges of that score. We believe that the proposed design and analysis plan are generally not valid.

Baker and Sargent (1) stated that “It may appear as though the control arm is used for both model fitting and treatment evaluation, a procedure that could lead to bias. However, this is not the case because the control arm provides no information on the effect of treatment.” Unfortunately, this is incorrect. For example, selecting a subgroup with the highest observed risk will tend to identify a subset of the control arm patients with poor observed outcome at least in part because of random outcome fluctuation (“random low”). Comparing this random low control arm subset with the corresponding experimental arm patient subset will result in bias (overestimation of treatment effect). This bias can be quite large as illustrated by the following simulations. In the simplest approach, we simulated two-arm RCTs with 100 patients per arm; normally distributed outcomes with both arms having the same population mean and variance were used; binary covariates taking value 0 or 1 with probabilities 0.5 for each patient independently of outcome and treatment were assumed. For each simulated RCT, the control arm data were used to build a simple one-covariate prognostic model using the covariate, which was most prognostic in the control arm based on a two-sided Student t test. Finally, we compared outcomes for the new treatment vs control arms in the poor-prognosis subgroup (defined by the selected covariate) using a two-sided P value of .025 as the threshold of statistical significance. In 5000 simulation replications, the type I error was inflated above the .025 nominal level, as shown in Table 1. We performed other simulations varying the types of covariates (eg, continuous), their covariance structure, and their prognostic effects; we also varied the modeling strategy and the number of patients but always with no true effect of treatment arm for any subgroup. Some of these results are also shown in Table 1. Although these factors influenced the size of the bias, in all cases the type I error was inflated. The size of the bias can be quite large unless there are a few highly prognostic covariates and a large number of patients.

In some cases (eg, adjuvant setting), it may be useful to focus evaluation of a new treatment in patients who do not have good prognosis on the control treatment.

Table 1. Observed type I error under no treatment effect for any subgroup in comparing outcome for new treatment and control arms in a poor prognosis subgroup determined by modeling the control arm (using a two-sided Student t test at the .025 nominal level)†

<table>
<thead>
<tr>
<th>No. of covariates</th>
<th>Binary covariates</th>
<th>Continuous covariates (select poor prognosis subgroup based on univariate linear regression)†</th>
<th>Continuous covariates (select poor prognosis subgroup based on optimal cut point)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>.08</td>
<td>.06</td>
<td>.18</td>
</tr>
<tr>
<td>100</td>
<td>.17</td>
<td>.12</td>
<td>.27</td>
</tr>
<tr>
<td>1000</td>
<td>.27</td>
<td>.21</td>
<td>.35</td>
</tr>
</tbody>
</table>

* Results are based on 5000 simulations of randomized clinical trials with 100 patients per arm.
† Univariate linear regression was used to model prognosis. The covariate with the smallest P value was determined, and its worst prognostic quartile was selected.
‡ For each covariate, all possible cut points for separating patients into poor- vs good-prognosis subgroups were examined, and the “optimal” cut point (separation) corresponding to the smallest P value (using two-sided Student t test) was determined. Then, the covariate with the smallest P value for the optimal separation was determined, and the corresponding worst prognosis subgroup was selected.

Prognostic models for such uses are often available before the start of the RCT but can be developed and used during the trial using the Adaptive Signature Design (2) and the Cross-validated Adaptive Signature Design (3). These designs are carefully structured to separate model development from model use for patient classification in comparing treatment arms. Using a prognostic model for control arm patients to adaptively define a subgroup for comparing treatments does not reduce the critical importance of maintaining the separation of model development and model validation that we emphasized.

RICHARD M. SIMON
BORIS FREIDLIN

References


Notes

Affiliation of authors: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD (RMS, BF).
Correspondence to: Richard M. Simon, DSc, Biometric Research Branch, National Cancer Institute, 9000 Rockville Pike, MSC 7434, Bethesda, MD 20892 (e-mail: rsimon@nih.gov).
DOI: 10.1093/jnci/djq552
Published by Oxford University Press 2011.
Advance Access publication on January 12, 2011.