Author's response to reviews

Title: Preference-adaptive randomization in comparative effectiveness studies

Authors:

Benjamin French (bcfrench@upenn.edu)
Dylan S Small (dsmall@wharton.upenn.edu)
Julie Novak (julnovak@wharton.upenn.edu)
Kathryn A Saulsgiver (kasau@upenn.edu)
Michael O Harhay (mharhay@upenn.edu)
David A Asch (asch@wharton.upenn.edu)
Kevin G Volpp (volpp70@exchange.upenn.edu)
Scott D Halpern (shalpern@exchange.upenn.edu)

Version: 2 Date: 8 January 2015

Author's response to reviews: see over
8 January 2015

Editorial Board
Trials

Dear Editors:

Thank you for the opportunity to revise and resubmit our manuscript, “Preference-adaptive randomization in comparative effectiveness studies” (MS: 4668675311469505). We carefully considered the comments from the referees and accordingly revised the manuscript. Below we include each referee’s comment in italic type, followed by our response in standard type. Note that reference and equation numbers herein correspond to those in the manuscript.

Regarding the editorial requests, we have included the date of registration with the trial registration number at the end of the Abstract (clinicaltrials.gov: NCT01526265; 31 January 2012), and have included a legend for Figure 1 after the reference list. Note that we used the BMC LaTeX style file to typeset our manuscript; the TeX file that we have supplied includes the affiliations and email address of all authors.

Thank you for considering our manuscript.

Sincerely,

[Signature]

Benjamin French, PhD
Assistant Professor of Biostatistics
Referee 1

This is a well-written paper describing a novel approach to account for differential acceptance of (or adherence to) assigned interventions. There are a few minor clarifications which would be nice to add prior to this paper being published.

We appreciate your positive review and the opportunity to clarify our approach.

In equations (2) and (3), the subscript i is used to indicate which stratum the subject is in, but in equation (4) i is used to indicate which intervention the subject is assigned to. It would be nice to be consistent, so consider changing “i” to “j” in equation (4).

Thank you. We have changed the subscript i to j in equation (4).

I don’t entirely understand what is meant on page 5 that “overly frequent updating would be inefficient because the procedure would overreact to trends predicated on small samples of new participants”. Please clarify this a bit more.

We apologize for the lack of clarity. Recall that the acceptance rates are calculated as the ratio of the number of participants who accept their intervention to the number of participants randomized. These rates could be highly variable, particularly early in the study, if the denominators used in the acceptance rates represent small samples of participants. We have modified the sentence (page 5):

We hypothesized that insufficiently frequent updating would enable certain arms to grow disproportionately before change was enforced. Conversely, we hypothesized that overly frequent updating would be inefficient because, particularly early in the study, the procedure would overreact to variable acceptance rates, for which the denominators represented small samples of participants.

In the illustration example, it sounds like subjects in the collaborative arms will have outcomes that are ultimately inevitably correlated with one another, perhaps necessitating the use of random effects in the final analyses to account for such non-independence. Please comment on whether the need for this type of analysis could or should be considered when updating the intervention assignment probability weights. This may be beyond the scope of this paper, but if the authors believe it is an area of future methodological development, perhaps they could just state so.

In our illustration, participants randomized to the collaborative reward or competitive deposit arm were assigned to a group after they accepted their assigned intervention. Because assignment to groups occurred after acceptance, there cannot be group effects on the interventions’ acceptance, and therefore no need to consider group effects when updating the allocation probabilities. In the Methods section of the Illustration (page 9), we clarify that assignment to groups occurred after acceptance:

Participants who accepted the collaborative reward or competitive deposit intervention were subsequently assigned to a group of six participants. Because group assignment occurred after acceptance, it was not necessary to adjust for group effects when calculating acceptance rates.
Based on the need to introduce the ceilings for allocation proportions, it seems in one sense that the authors developed an algorithm that ultimately proved to be impractical. I think they need to discuss this a bit more in the discussion. For example, should their algorithm be updated in the future to allow for such modifications? How might this impact their assessments of efficiency?

We agree that discussion regarding automation of our procedure is necessary. We provide such discussion on page 13:

Second, in our application the procedure required manual modification due to lower-than-anticipated acceptance of less appealing arms, and correspondingly higher-than-anticipated automatic adjustments to the allocation probabilities to those arms. Left unchecked, those automatic adjustments would have hampered our ability to adequately enroll any of the arms. Future investigators may wish to program automatic modifications in their preference-adaptive randomization procedure, similar to the modifications that we made manually (e.g., ceilings for the allocation probabilities).
Referee 2

This paper is very well written and easy to read. It exposes a novel method of implementing randomization in case we may suspect a high refusal rate of the randomization result among patients. The strength of the approach is enforced by a real-world example.

Thank you for your positive review of our work.

I however feel a bit disappointed by the fact that authors only focused on the randomization step, without illustrating the implication of their approach on the treatment effect estimation. Thus they clearly demonstrated that using their approach, they were able to obtain groups of patients with acceptable balance (in numbers). They then evoked using an instrumental variable approach to analyze data, but without providing any result. I would indeed be much more interested in having a complete view of the issue, which would be more convincing. Indeed, a numerical balance between groups is of importance for power, but a non-biased assessment of the intervention effect is surely of greater importance. My feeling is therefore that we miss a section demonstrating that such an approach is able to provide non biased treatment effect estimates.

We agree that additional details regarding treatment effect estimation would be useful. Methods based on instrumental variables are well studied in the context of providing unbiased treatment effect estimates in the presence of selection effects. Therefore, we have moved the discussion regarding instrumental variables to a new section before the Discussion (entitled Estimation of Treatment Effects), expanded on its presentation, and provided additional references (page 11-12):

In previous sections, we described the development and implementation of a preference-adaptive randomization procedure in comparative effectiveness studies, using a smoking cessation trial as an illustrative example. In this section, we outline an analysis plan to estimate acceptance, efficacy, and effectiveness; see equation (1).

A standard intention-to-treat analysis will be used to compare the treatments' acceptance and effectiveness. For evaluating efficacy, a standard per-protocol analysis would compare the cessation rate among participants who accept intervention j to those who receive the control (with or without inclusion of those who were assigned intervention j but declined it). However, the standard per-protocol analysis may be subject to selection bias if smokers who do not accept an incentive differ from those who do in ways that relate to their probabilities of quitting [23]. To address such selection effects, we will model the randomization arm as an instrumental variable [11]. In the instrumental variable approach, the cessation rate of each intervention is adjusted for the percentage of participants who accept their assigned intervention, thereby estimating complier-averaged causal effects and attenuating the selection effects [24, 25].

A key advantage of the instrumental variable approach over a standard per-protocol analysis is that the instrumental variable approach uses the data on all randomized participants, rather than merely those who accept their assigned intervention. Therefore, the instrumental variable approach adheres to the randomized trial principle that participants should be analyzed according to their randomization status, rather than according to their self-selected acceptance status. For the instrumental variable analysis, we will use a two-stage least squares linear probability model [26]. By using the preference-adaptive randomization
procedure that balances the number of accepting participants in each arm (or, in our application, the arms targeted for complete enrollment), we increase the power for the instrumental variable analysis.

For our Illustration, the primary efficacy results were not available at the time we wrote this paper. Those results have been included in the main trial paper, which is currently under review in a clinical journal.

Table 4 displays the characteristics of the patients randomized and, unsurprisingly, there is no imbalance. Indeed, I don’t see why changing the randomization ratios over the trial would induce imbalance. However, authors specified (in the Discussion section!) that “balance across arms was not uniformly achieved when evaluating only participants who accepted their assigned intervention” (which is not surprising, since acceptance is probably surely linked to some baseline characteristics, which may also be prognostic factors), and I indeed would be very interested in seeing these imbalances. Authors cannot just say “data not shown”.

Changing the randomization ratios over time could result in imbalances if participant characteristics were not stable over time. For example, in our Illustration, various recruitment methods were rolled out at different times during the study period, such as mailed postcards, emails, and messages on paychecks. These methods could have differed in the type of participants that they recruited, and therefore the characteristics of recruited participants could have changed over time.

We agree that showing participants’ characteristics according to their accepted intervention is of interest, which we have done in a new Table 4. The results are presented in the Results section of the Illustration (page 10):

Balance across arms was not uniformly achieved when evaluating only participants who accepted their assigned intervention (Table 4). In particular, annual household income was highly imbalanced, with an over-representation of high-income individuals in the individual deposit and competitive deposit arms ($p < 0.001$). These results provide support for the concern that analyses based on participants who accepted their assigned intervention would be susceptible to selection effects. In the following section, we discuss instrumental variable methods that can address such selection effects.

The proposed approach supposes that there is no participant blinding in the trial, and that we may suspect a differential acceptance rate. Such a context is not so frequent, although authors indeed illustrate it with a real-world example. I think some specifications of the context in which such an approach could be of help is necessary.

In our Illustration, participants were not blinded to the intervention. However, preference-adaptive randomization could be useful in blinded trials. In addition, acceptance (or more generally, adherence) is often an important issue in randomized trials. In the Discussion, we summarize the types of studies in which preference-adaptive randomization could be used. Examples include large simple trials of vaccines or virtually any pharmaceutical for which adherence may not be 100%. In such trials, which are typically blinded, a measure of adherence could be used to update the allocation probabilities. We now specifically state in the Discussion (page 13) that our approach could be applicable in both blinded and unblinded trials:

Our adaptive design, coupled with appropriate statistical analysis methods, could be used to enhance the validity and generalizability of any comparative effectiveness study, blinded or unblinded, in which study participants choose to adhere to their assigned intervention. Examples include large simple trials of vaccines or virtually any pharmaceutical for which adherence may not be 100%, and, of course, trials of almost any behavioral intervention.
Of course, a member of the study team would likely need to be unblinded (either fully or partially) in order to implement and monitor preference-adaptive randomization. An unblinded statistician is commonplace in randomized trials with adaptive designs or interim analyses. We have added this point to the Discussion (page 13):

Fourth, like many adaptive designs or interim analyses, preference-adaptive randomization might require that a member of the study team, such as a statistician, be unblinded during the trial. Unblinding requires careful consideration of the statistician’s role in the study’s conduct and reporting.