Supplementary Materials for "Blocking for Sequential Political Experiments"*

Ryan T. Moore† Sally A. Moore‡

23 April 2013 at 13:19

Abstract

This document accompanies Moore and Moore (2013a). Here, we give an example of our interactive data collection interface, show example calculations, and provide additional figures to supplement the primary paper.

*This work is supported in part by a Pilot Grant [UL1RR025014] from the Institute for Translational Health Sciences (ITHS). ITHS is funded by the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. This study was partially supported by the Office of Academic Affiliations, Advanced Fellowship Program in Mental Illness Research and Treatment, Department of Veterans Affairs, and by the Robert Wood Johnson Foundation. We thank Nicholas Beauchamp, Jens Hainmueller, Kosuke Imai, Rebecca Morton, Kevin Quinn, and the participants in EGAP 6 and ICHPS 9 for helpful suggestions. The replication archive is available as Moore and Moore (2013b).

†Department of Political Science, Washington University in St. Louis, 241 Seigle Hall, Campus Box 1063, One Brookings Drive, St. Louis MO 63130. tel: 314.935.3461; fax: 801.922.7901; `rtm` (at) `wustl` (dot) `edu`; `http://rtm.wustl.edu`.

‡VA Puget Sound Health Care System - Seattle Division, Evidence Based Treatment Centers of Seattle, and the University of Washington Department of Psychiatry and Behavioral Sciences.

1 Data Collection Query and Assignment

In sequential experiments, subject intake and treatment assignment is often done by research assistants or others. Thus, we offer a simple, interactive interface to allow the researcher conducting subject intake to enter participant $n + 1$'s data, which is then appended to a larger dataset of participants $1, \ldots, n$.

In our interface, the first participant's query requests data about the structure of the background data to be collected throughout the experiment, and thus includes many more questions than the query for $2, \ldots, N$. Implementation is in R (R Core Team, 2013).

An abbreviated sample query for the first participant of an experiment follows, with the software's query in black and the researcher's responses in red.

```
> seqblock1(query=T)
How many identification variables are there?
1
Enter the name of ID variable 1 without quotation marks.
id
Enter the value of 'id'.
10624
How many exact blocking variables are there?
0
How many blocking variables are there?
2
Enter the name of blocking variable 1.
x1
Enter the value of 'x1'.
100
Should 'x1' be restricted to certain values?  [n/y]
no
Enter the name of blocking variable 2.
x2
Enter the value of 'x2'.
80
How many experimental/treatment conditions are there?
2
```

Other arguments queried include condition names, the output file name, variable data types, initial assignment probabilities, a random seed, the statistic used to summarize the current

subject's covariate relationship to the already-assigned subjects' profiles, and the method for setting the π_t assignment probabilities.

To allow the researcher to immediately implement a treatment, the new participant's treatment assignment is displayed. Further, as a double-check and to ensure transparency, the interface provides a printout of the participant's data as entered and the location of the current working directory.

```
Unit 1 data stored as file sbout1.RData.
The current working directory is /Users/you/Documents/yourdir
Unit 1 assigned to Treatment 1.
The new data as entered:
    id   x1  x2    Tr Assg
 10624  100  80  Treatment 1
```

When participants after the first one enter the study, metadata such as the variable names and types is not requested, so the query is shorter. The first step is to provide the already-assigned data:

```
> seqblock2k(query=T)
Enter the name of the input file without quotation marks.  [E.g., sbout1.RData]
sbout1.RData
Enter the value of 'id'.
8333
Enter the value of 'x1'.
90
Enter the value of 'x2'.
98
Unit 1:2 data stored as file sbout2k.RData.
The current working directory is /Users/you/Documents/yourdir
Unit 2 assigned to Treatment 2.
The new data as entered:
    id x1 x2 Tr
2 8333 90 98 Treatment 2
```

2 Supplementary Calculations and Figures

2.1 Example Calculations for Section 3

Below we show example calculations for assignment mappings that most directly employ the distribution of Mahalanobis distances between the current unit and the units in each treatment condition.

A first method assigns the current unit to treatment t with probability $\frac{\overline{MD}_{qt}}{\sum_{t=1}^{T} \overline{MD}_{qt}}$. To illustrate, suppose there are three treatment conditions: treatment A has already been assigned to units with MD's from the current unit of 1, 2, 3, 4, and 5; treatment B to units with MD's of 1 and 2, and C to units with MD's of 1, 2, and 3. Then, the assignment probabilities would be $\frac{3}{3+1.5+2}$, $\frac{1.5}{3+1.5+2}$, and $\frac{2}{3+1.5+2}$.

A second method assigns conditions proportionally to the sum of the Mahalanobis distances between that condition and the current unit. Specifically, condition t is assigned probability $\sum_{r=1}^{R} MD_{qr}$, normalized by the total $\sum_{t=1}^{T}\sum_{r=1}^{R} MD_{qr}$. Using the same example distributions of MDs by treatment condition as above, the assignment probabilities are $\frac{15}{24}$, $\frac{3}{24}$, and $\frac{6}{24}$.

A further alternative uses the squares of the sums of the MDs, yielding example probabilities $\frac{15^2}{15^2+3^2+6^2}$, $\frac{3^2}{15^2+3^2+6^2}$, and $\frac{6^2}{15^2+3^2+6^2}$.

2.2 Supplementary Figures

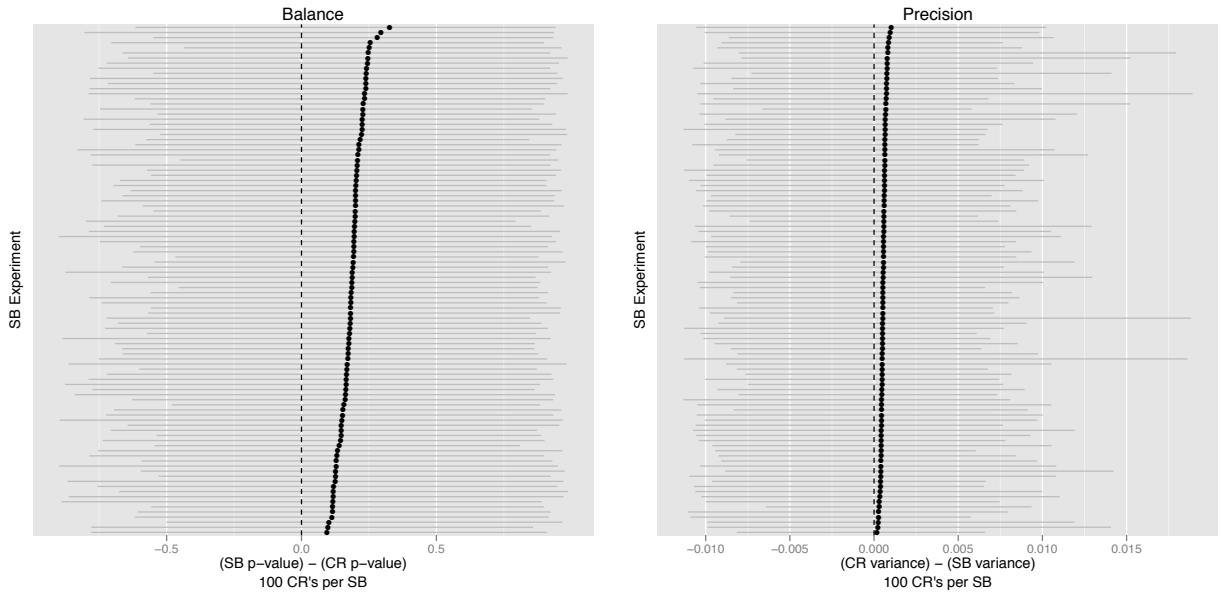

Figure 1: Sequential blocking outperforms complete randomization in balance and precision. One hundred blocked experiments are simulated, each completely rerandomized 100 times. Uncorrelated MVN data; same 100 sets of covariates represented in each segment. Different segments represent different rerandomizations. See Section 4.1.

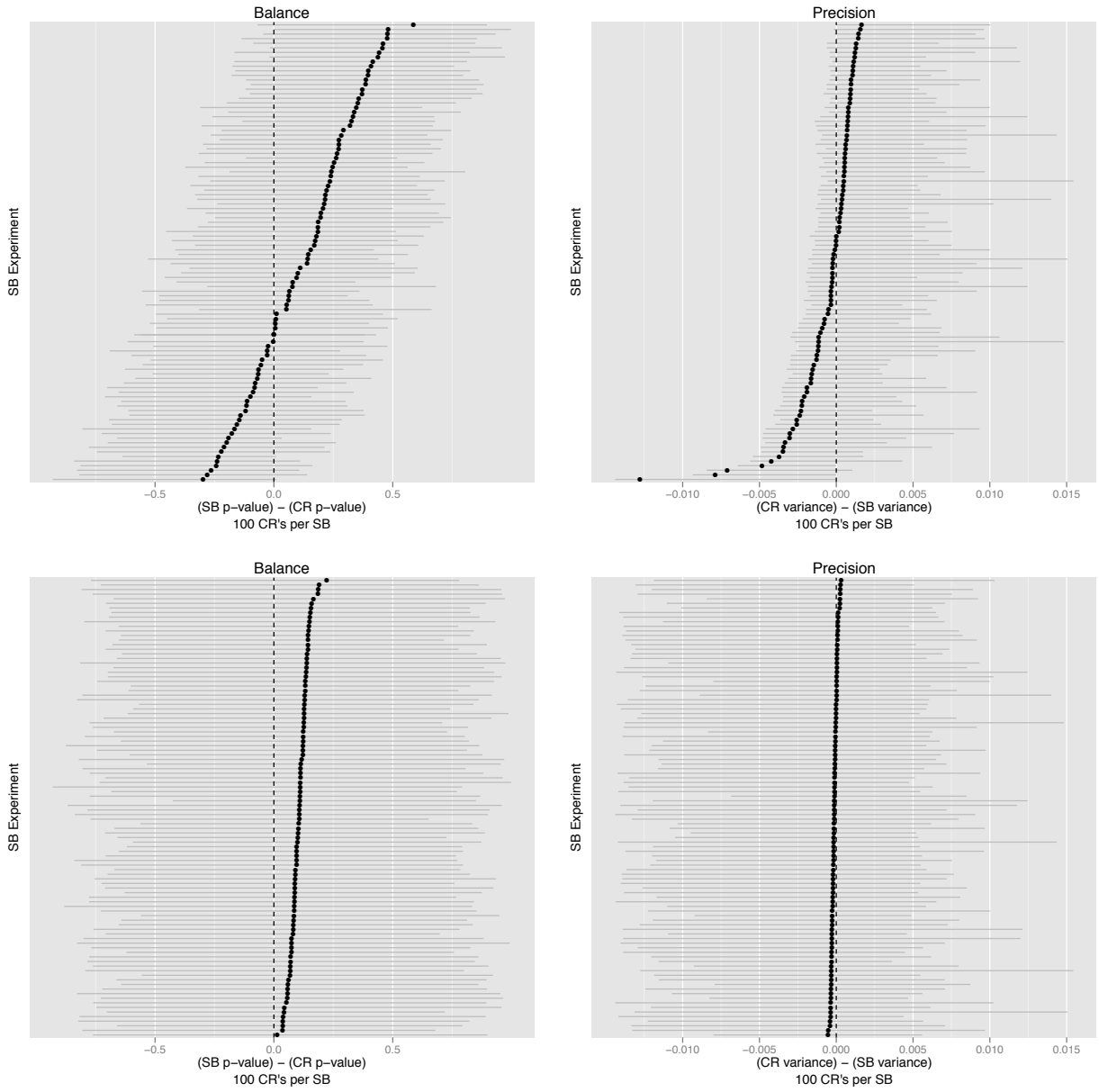

Figure 2: Sequential blocking outperforms complete randomization in balance and precision. One hundred blocked experiments are simulated, each completely rerandomized 100 times. MVN data correlated at $r = 0.6$, with extreme outlier introduced at unit 20. Top: each segment represents one set of covariates, one SB minus 100 CR's. Bottom: each segment represents 100 sets of covariates, 100 SB's minus 100 CR's. See Section 4.3.

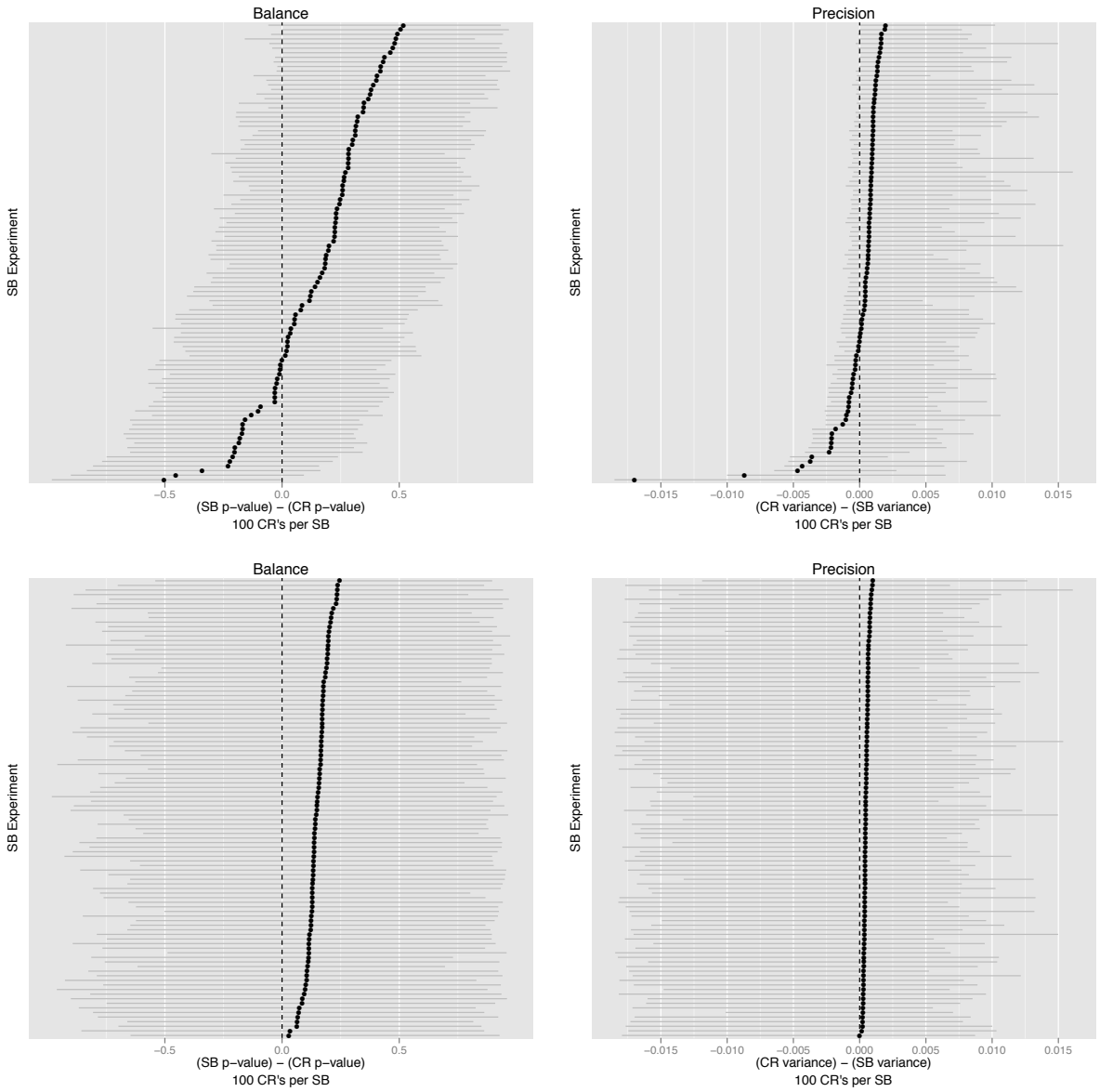

Figure 3: SB experiments more balanced and precise than CR experiments, aggregating two ways. One hundred blocked experiments are simulated, each completely rerandomized 100 times. In all panels, values to the right represent sequential blocking's advantage over complete randomization. At top, a segment represents range of differences for one SB minus 100 CR's; at bottom, a segment represents range of differences for 100 SB's minus 100 CR's; points represent median differences. Highly correlated bimodal MVN data $r = 0.8$; see Section 4.4.

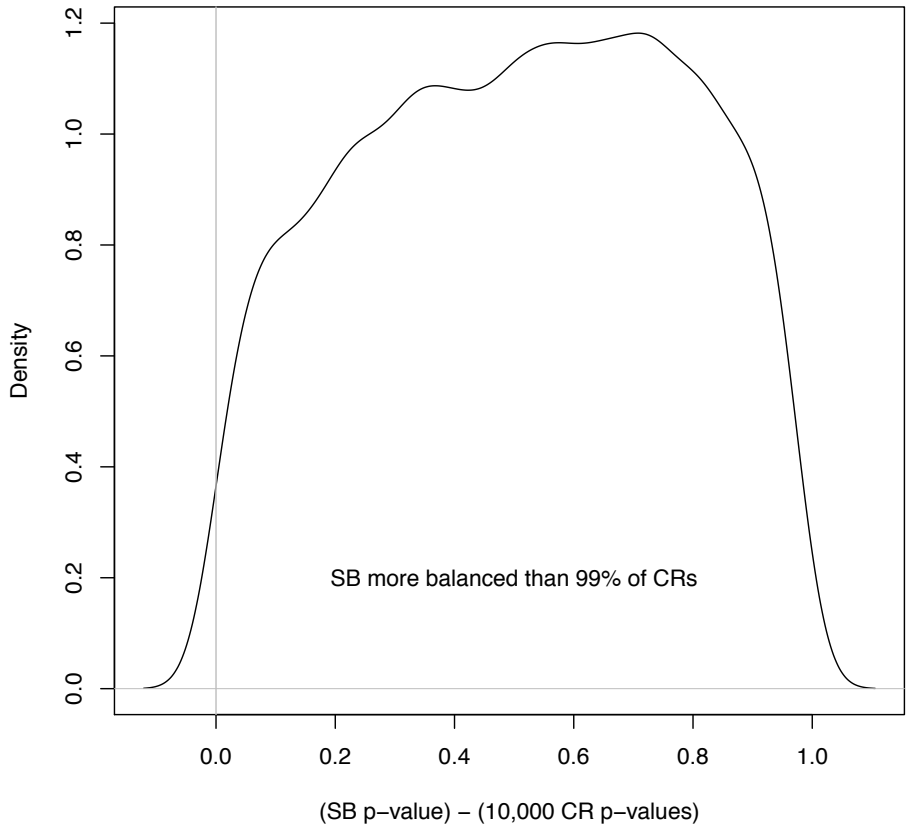

Figure 4: Sequentially blocked PTSD trial is more balanced than ninety-nine percent of 10,000 complete rerandomizations of the observed trial.

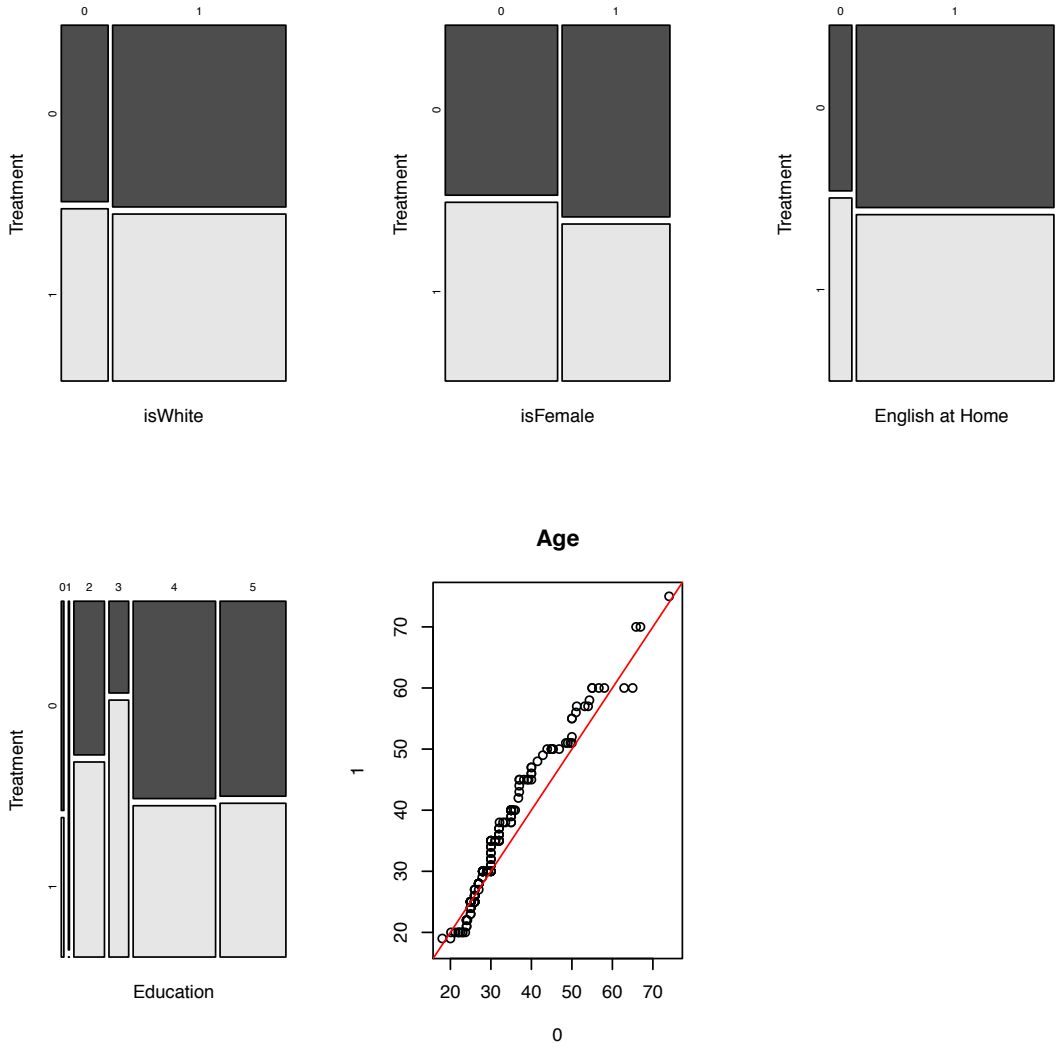

Figure 5: Balance in the Cobb, Greiner and Quinn (2011) data subset. Mosaic and quantile-quantile plots for four discrete and one quasi-continuous covariates reveal imbalance in age and education.

References

Cobb, Rachael V., D. James Greiner and Kevin M. Quinn. 2011. "Can Voter ID Laws be Administered in a Race-Neutral Manner? Evidence from the City of Boston in 2008." *Quarterly Journal of Political Science* 6:1–33.

Moore, Ryan T. and Sally A. Moore. 2013a. "Blocking for Sequential Political Experiments." *Political Analysis*. Forthcoming.

Moore, Ryan T. and Sally A. Moore. 2013b. "Replication data for: Blocking for Sequential Political Experiments." http://hdl.handle.net/1902.1/21042. IQSS Dataverse Network, V1.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.