A Priori Matching in Cluster Randomized Trials

true

April 28, 2017

Introduction

To determine the efficacy of a treatment, individually randomized trials (IRTs) with blinding are the "strongest study design available" (**?**). Unfortunately, cost and study design amongst other things mean some interventions can not be randomized on an individual level. For example, training medical doctors to improve patient interactions would require individuals be randomized to a doctor not of their choosing and assumes no medical emergencies treated by other doctors for the duration of the study. To remedy this, investigators could consider randomizing medical practices to intervention or control groups. Trials where groups of participants are randomized are called cluster randomized trials (CRTs). Three reasons for conducting a CRTs are: (i) implementation occurs at the cluster level, (ii) to avoid contamination, and (iii) to measure intervention effects among cluster members who do not receive treatment (**??**).

Many authors discuss the importance of matching in CRTs (**????????**). SUMMARY HERE

Despite all this information few authors discuss methodologies to support the matching process. We suggest a method suitable for *a priori* matching using baseline data. In section 2 we outline our method, section 3 applies it to the PROTECT dataset, and section 4 is a brief discussion.

Methods

To approach this complex topic of balancing randomization in CRTs we suggest a new approach. Our approach involves weighting variables of import, matching units using these weights, and randomizing many times to obtain a distribution of possiblities when official randomization occurs. Investigators assess these distributions to determine if possible randomizations are sufficiently balanced, if not, weighting is adjusted and the process begins again. The details follow.

The initial step involves prioritizing variables $(1, 2, ..., m)$ from units $(1, 2, ..., n)$ to be randomized. We have

$$
\begin{aligned}
\overline{V_1} &= (v_{11}, v_{12}, ..., v_{1n}) \\
\overline{V_2} &= (v_{21}, v_{22}, ..., v_{2n}) \\
\vdots &= \vdots \\
\overline{V_m} &= (v_{m1}, v_{m2}, ..., v_{mn}).
\end{aligned}
$$

In addition, we use $\overline{w} = (w_1, w_2, ..., w_m)$ to weight and standardize $(\overline{V_1}, \overline{V_2}, ..., \overline{V_m})$. We have NOT SURE IF THIS IS RIGHT, THINK IT'S NOT!!!!!!!! something with ij is wrong.

$$
v_{ii}^* = \frac{(v_{ii} - \frac{\sum_{j=1}^{n} v_{ij}}{n}) * w_i}{sd(\overline{V_i})}
$$

where $sd(\overline{V_i})$ is the standard deviation of $\overline{V_i}$. We now have

$$
\begin{aligned}
\overline{V_1^*} &= (v_{11}^*, v_{12}^*, ..., v_{1n}^*) \\
\overline{V_2^*} &= (v_{21}^*, v_{22}^*, ..., v_{2n}^*) \\
\vdots &= \vdots \\
\overline{V_m^*} &= (v_{m1}^*, v_{m2}^*, ..., v_{mn}^*).
\end{aligned}
$$

which we use to compute the Mahalanobois Distance matrix, $\mathbf{D}$. From here we use the `nmatch` function in the `designmatch` package in `R` to find $\frac{n}{2}$ pairs if n is even. If n is odd, the remainder can be randomized to treatment or control per the principal investigator. Without loss of generality, we assume n is even for the remainder of this paper and note that to include an odd n either treatment or control groups will include one more set of priority variables.

Once the matching is completed and pairs found we return to using the raw data, as this will be used to assess the weighting scheme. We now have pairs $(\overline{V}_{11}, \overline{V}_{12}), (\overline{V}_{21}, \overline{V}_{22}), ..., (\overline{V}_{\frac{n}{2}1}, \overline{V}_{\frac{n}{2}2})$. The first match in each pair will be randomized to either treatment or control using the `rbinom` function in `R`. Next, we subset $\overline{V}_1, \overline{V}_2, ..., \overline{V}_m$ into appropriate randomization subgroups: $\overline{V}_{1T}, \overline{V}_{1C}, \overline{V}_{2T}, \overline{V}_{2C}, ..., \overline{V}_{\frac{n}{2}T}, \overline{V}_{\frac{n}{2}C}$ where $\overline{V}_{iT} = (v_{i1}^T, v_{i2}^T, ..., v_{in}^T)$, similarly for $\overline{V}_{iC}$. Using these we find

$$
k_j = | \sum_{i=1}^{\frac{n}{2}} v_{ij}^T - \sum_{i=1}^{\frac{n}{2}} v_{ij}^C |
$$

THE ABOVE NEEDS A BETTER ABSOLUTE VALUE SYMBOL. for $j = 1, 2, ..., m$. We randomize N times and find k_{lj} the difference in the two arms for the jth priority variable for each of the $l = 1, 2, ..., N$ re-randomizations. To assist analysis we draw a parallel coordinates plot where the jth axis plots k_{lj} for $l = 1, 2, ..., N$. If the principal investigator finds the possible differences too large for a priority variable j, increasing w_j and re-running the above will update the matching to attain closer matches for this variable and lessen the differences. The penality in this process is that closer matches for variable j are likely to imply reduced closeness in another variable, so compromises must be made.

Results

To demonstrate the usefulness of this technique we present a brief summary of our randomization process using baseline data from the PROTECT trial (Project PROTECT: Protecting Nursing Homes From Infections and Hospitalization) (**?**). In this trial, the investigators are studying whether bathing with chlorhexidine gluconate and iodophor nasal swabs "can reduce hospitalizations associated with infections, antibiotic utilization, and multi-drug resistant organism (MDRO) prevalence" versus regular bathing. Additional training is given to nursing homes in the treatment arm to ensure comprehension and adherence to trial protocol.

Prior randomize baseline data was collected for ?????? on the 29???? nursing homes. With this data, investigators met to prioritize variables into several categories: primary, secondary, tertiary, and not relevant. For this trial, the investigators decided that percentage discharges to hospital with infection based on primary and other diagnsoses, percentage discharge to hospital, percentages MDRO, and percentage usage of antibacterials started at nursing home were of primary importance. Of secondary importance were percentage of admissions with length of stay over 100 days, average daily census, and mean number of baths per resident per week. If they were able to include more variables without effecting the balance of the others they felt matching on AVERAGE DEPENDENT late LOSS activities of daily lives, and the Centers for Medicare and Medicaid Services (CMS) rating of eaching nursing home. To ease understanding, our initial discussion will involve 2 variables: percentage discharges to hospital with infection based on primary and other diagnsoses, and average daily census.

Table 1. Abbreviations of variables used to randomize

Primary import	Secondary import	Tetiary import
% Inf	% Long Stay	Late ADLs
% DC	Avg Daily Census	CMS Star
MDRO	Baths/Week	
% Abx		

Prior to randomization, investigators

Discussion

Look how smart we are.

References

Laura B Balzer, Maya L Petersen, and Mark J van der Laan. Why match in individually and cluster randomized trials? 2012.

Elizabeth DeLong, Lingling Li, and Andrea Cook. Pair-matching vs stratification in cluster-randomized trials. URL https://www.nihcollaboratory.org/Products/Pairing-vs-stratification_V1.0.pdf.

Allan Donner, Monica Taljaard, and Neil Klar. The merits of breaking the matches: a cautionary tale. *Statistics in medicine*, 26(9):2036–2051, 2007.

Constantine Gatsonis and Sally C Morton. Methods in comparative effectiveness research, 2017.

Moulton Hayes. *Cluster Randomised Trials.* Chapman and HallCRC, 2009.

Kosuke Imai, Gary King, Clayton Nall, et al. The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science*, 24 (1):29–53, 2009.

Neil Klar and Allan Donner. The merits of matching in community intervention trials: a cautionary tale. *Statistics in medicine*, 16(15):1753–1764, 1997.

D.M. Murray. *Design and Analysis of Group-randomized Trials.* Number v. 29; v. 1998 in Design and Analysis of Group-randomized Trials. Oxford University Press, 1998. ISBN 9780195120363. URL https://books.google.com/books?id=cVLs3m4a9ZoC.