Matching in Cluster Randomized Trials Using the Glodilocks Approach

true

May 19, 2017

Introduction

To determine the efficacy of a treatment, individually randomized trials (IRTs) with blinding are the "strongest study design available" [Gatsonis and Morton, 2017]. Unfortunately, cost and study design amongst other things mean some interventions can not be randomized on an individual level. For example, education researchers decide to determine if training elementary school teachers in a reading program will affect literacy skills in third graders. Randomizing each third grader to treatment or control would not be suitable, as this equates to randomly allocating students to a teacher; rural schools tend to be smaller with only one teacher which further complicates matters. Instead, researchers may choose to randomize teachers, schools, or counties to determine any differences between the two arms of the study. Trials where groups 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 [Balzer et al., 2012, Hayes, 2009]. CRTs are "the gold standard when allocation of identifiable groups is necessary" [Murray et al., 2004].

DO I NEED TO SAY SOMETHING ABOUT MATCHING HERE?? PROBABLY SO??

Many authors debate matching in CRTs [Balzer et al., 2012, Hayes, 2009, Gatsonis and Morton, 2017, Diehr et al., 1995, Murray, 1998, Imai et al., 2009, DeLong et al., Donner et al., 2007, Klar and Donner, 1997, Donner and Klar, 2000, Martin et al., 1993]. Murray argues that "the choice of matching or stratification [of] factors is critical to the success of the procedure" [Murray, 1998]. Some agree that caution must be used when matching a small number of clusters due to the decrease in power [Donner and Klar, 2000, Klar and Donner, 1997, Balzer et al., 2012, Martin et al., 1993]. Breaking matches in the analysis stage addresses this [Diehr et al., 1995], but perhaps only when there are a small number of large clusters [Donner et al., 2007]. Others argue that matching is effective in a small number of clusters as it "increases the chance of the intervention groups being well-balanced" [Donner et al., 2007]. Imai et al argue that not matching, in small or large sample, is "equivalent to discarding a considerable fraction of one's data" [Imai et al., 2009]. However, in one trial "matching actually led to a loss in statistical efficiency" [Manun'ebo et al., 1994, Donner and Klar [2000]].

Despite all this debate few authors discuss methodologies to support the matching process [Raab and Butcher, 2001]. An alternative to matching suggests balancing criterion based on conditioned means of covariates [Raab and Butcher, 2001]. Our article is an extension of methods introduced in Chapter 4 of Methods in Comparative Effectiveness Research [Gatsonis and Morton, 2017]. 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 SWAPOUT 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 possibilities 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

$$v^*_{ij} = \frac{(v_{ij} - \frac{\sum_{k=1}^{n} v_{ik}}{n}) * w_i}{sd(\overline{V_i})}$$

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` [Greevy et al., 2004] package in `R` [Zubizarreta and Kilcioglu]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 $\overline{V}_{iT} = (v^T_{i1}, v^T_{i2}, ..., v^T_{in})$, similarly for $\overline{V}_{iC}$. Using these we find

$$k_j = |\sum_{i=1}^{\frac{n}{2}} v^T_{ij} - \sum_{i=1}^{\frac{n}{2}} v^C_{ij}|$$

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 j^{th} priority variable for each of the $l = 1, 2, ..., N$ re-randomizations. To assist analysis we draw a parallel coordinates plot where the j^{th} 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 SWAPOUT trial (Cluster-randomized Non-inferiority Trial Comparing Mupirocin vs Iodophor for Nasal Decolonization of ICU Patients to Assess Impact on Staphylococcus aureus Clinical Cultures and All-cause Bloodstream Infection During Routine Chlorhexidine Bathing) [Platt]. In this non-inferiority trial, the investigators are studying whether bathing with chlorhexidine gluconate and swabbing iodophor nasal swabs are inferior to bathing with the same and mupirocin nasal swabs. REDUCE trial [Huang et al., 2013] mupirocin nasal swabs and bathing with chlorhexidine reduced the MRSA Staphylococcus aureus (an antibiotic resistant infection) in Hospital Corporation of America intensive care units (ICU). However, physicians are reluctant to use mupirocin, an antibiotic, so investigators are swapping it with iodophor.

Table 1. Abbreviations of variables used to randomize

Primary	Secondary	Tetiary	Quaternary	Quinary
Pt Days	Median LOS	Medicaid	DC SNF	Onc_BMT_Trp
S auer Rate	Comorbidity Score	PCR Blood	Surgery	BMT_Trp
MRSA Rate				
All Blood				
Mup-R				
Hx MRSA				
Mup Adherence				
CHG Adherence				

Prior to randomization baseline data was collected for 20 months on the 137 hospitals. With this data, investigators met to prioritize baseline variables into several categories: primary, secondary, tertiary, quaternary, quinary, and not relevant to randomization. For this trial, the investigators decided that average monthly attributable days, Staphylococcus aureus Intensive Care Unit (ICU)-attributable cultures per 1,000 days, MRSA ICU-attributable cultures per 1,000 days, all pathogen ICU-attributable bacteremia cultures per 1,000 days, regional mupirocin resistance estimate, percent of admissions with MRSA diagnosis within a year, percent of mupirocin use admission to day 5, survey Chlorhexidine Glucominate were all of primary importance. Of secondary importance were median ICU length of stay, and mean elixhauser total score. Of tertiary importance were the percentage of ICU medicaid patients, and whether or not a facility uses Polymerase chain reactions to identify MRSA in blood. Next, percent admissions to skilled nursing facility (SNF), and the percent of admissions with Center for Disease Control and Prevention surveillance surgery. Lastly, if the ICU has specialty units for oncology, bone marrow transplant, or transplant units, and if the ICU has bone marrow transplant or transplant units. More information on each variable is avaliable in appendix 1 and their abbreviations, in the same order, can be found in table 1. To ease understanding, our initial discussion will involve the first 3 variables above: Patient days, Staphylococcus aureus rate, and MRSA rate.

Prior to randomization, investigators spent time using a web application built using the `Shiny` package in `R`. The purpose of this is to determine which weights give advantageous balance across relevant baseline variables. We recommend deciding on ideal and acheivable maximum differences in study arms and using many combinations of weights until one is found which ensures randomization is likely to be within those bounds. In the well-known childrens fable The Three Bears, Goldilocks tries three bowls of porridge, one is too hot, the other too cold, and the third is just right [3Be]. We recommend a similar procedure applied to weights, with perhaps more attempts.

Figure 1 demonstrates this process using three variables: attributable patient days per month, Staphylococcus auer rate, and MRSA rate. After initial explorations on the web application, investigators agreed that an ideal maximum mean differences in treatment and control arms for these variables were: 80 attributable patient days per month, 15% difference in Staphylococcus aureus infection rates, and 15% difference in MRSA rate. The graph on the top left shows no weighting on any of these variables, the maximums are

not consistently attained on any axis: there is a reasonable chance that if randomization occured with this weighting all variables would be above the desired maximum mean differences in treatment and control arms. To rectify this, nonzero weighting must be added. In the top-right graph a weight of 8 has been applied to the Staphylococcus auer rate. In this graph, the matching of hospitals is strongly skewed so that hospitals with similar Staphylococcus auer rates are paired. This results in low mean difference between the treatment and control arms in that variable. The middle axis is consistently below the maximum value: if randomization occured using these weights we are likely to get suitable balance in this variable. Unfortunately, there is a penalty. Hospitals with similar Staphylococcus auer rates may not have similar attributable patient days per month and MRSA rates, which results in increases in these values. In particular, our investigators felt that the chance of attaining MRSA rates above 15% were too high for this weighting. The bottom plot shows the possible mean balances used in the actual randomization for these three variables, the weights for each variable were 1, 4, and 2, respectively. In all three graphs, the black line indicates the mean value of all points on each axis.

Our investigators used this approach with 16 variables. After trying many weights this weighting provided the best balance between treatment and control arms for the variables of important. When the trial was randomized we used this weighting to match hospitals in the study, then randomized the first member of each match to either treatment control.

Discussion

While the Goldilocks approach to randomizing does not ensure balance in the treatment and control arms, it is a tool that provides investigators with a method to explore weights that impact matching and balance. We encourage investigators that utilize CRTs to use this method prior to randomizing to find more balance in treatment and control arms and SOMETHING THEORETICAL HERE!!!!

Future work in this area includes publishing a `Shiny` web application for investigators to utilise. This application will eventually be an interactive plot that enables users to click on each axis and view where low and high draws of that variable fall for other variables. In some cases, our investigators find that matching on 1 variable seems to give suitable balance throughout.

Appendix

A more formal explanation of the variables here, in table format, to be checked with Susan.

Variable	Description
Pt Days	
S auer Rate	
MRSA Rate	
All Blood	
Mup-R	
Hx MRSA	
Mup Adherence	
CHG Adherence	
Median LOS	
Medicaid	
Comorbidity Score	
Medicaid	
PCR Blood	
DC SNF	
Surgery	
Onc_BMT_Trp	

Variable	Description
BMT_Trp	

Table 2: Thorough description of baseline variables used in this paper.

References

URL https://en.wikipedia.org/wiki/Goldilocks_and_the_Three_Bears.

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.

Paula Diehr, Donald C Martin, Thomas Koepsell, and Allen Cheadle. Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Statistics in medicine*, 14(13):1491–1504, 1995.

A. Donner and N. Klar. *Design and Analysis of Cluster Randomization Trials in Health Research*. Wiley, 2000. ISBN 9780340691533. URL https://books.google.com/books?id=QJZrQgAACAAJ.

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.

Robert Greevy, Bo Lu, Jeffrey H. Silber, and Paul Rosenbaum. Optimal multivariate matching before randomization. *Biostatistics*, 5(2):263, 2004. doi: 10.1093/biostatistics/5.2.263. URL +http://dx.doi.org/10.1093/biostatistics/5.2.263.

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

Susan S Huang, Edward Septimus, Ken Kleinman, Julia Moody, Jason Hickok, Taliser R Avery, Julie Lankiewicz, Adrijana Gombosev, Leah Terpstra, Fallon Hartford, et al. Targeted versus universal decolonization to prevent icu infection. *New England Journal of Medicine*, 368(24):2255–2265, 2013.

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.

Manwela N Manun'ebo, Patricia A Haggerty, Muladi Kalen Gaie, Ann Ashworth, and Betty R Kirkwood. Influence of demographic, socioeconomic and. *Journal of tropical medicine and hygiene*, 97:31–38, 1994.

Donald C Martin, Paula Diehr, Edward B Perrin, and Thomas D Koepsell. The effect of matching on the power of randomized community intervention studies. *Statistics in medicine*, 12(3-4):329–338, 1993.

David M Murray, Sherri P Varnell, and Jonathan L Blitstein. Design and analysis of group-randomized trials: a review of recent methodological developments. *American journal of public health*, 94(3):423–432, 2004.

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.

Richard Platt. Mupirocin-iodophor icu decolonization swap out trial. URL https://clinicaltrials.gov/ct2/show/NCT03140423?term=swap+out&rank=1.

Gillian M Raab and Izzy Butcher. Balance in cluster randomized trials. *Statistics in medicine*, 20(3):351–365, 2001.

J. R. Zubizarreta and C. Kilcioglu. designmatch: Construction of optimally matched samples for randomized experiments and observational studies that are balanced and representative by design.