

Matching in Cluster Randomized Trials

true

April 18, 2017

Introduction

To determine the efficacy of a treatment, individually randomized trials (IRTS) are the gold standard.

Scholarship: How are people matching CRTs? Read and find out.

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 $(1, 2, \dots, m)$ from units $(1, 2, \dots, n)$ to be randomized. We have

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

In addition, we use $\overline{w} = (w_1, w_2, \dots, w_m)$ to weight and standardize $(\overline{V_1}, \overline{V_2}, \dots, \overline{V_m})$. 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

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

which we use to compute the Mahalanobis Distance matrix, **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 $(\bar{V}_{11}, \bar{V}_{12}), (\bar{V}_{21}, \bar{V}_{22}), \dots, (\bar{V}_{\frac{n}{2}1}, \bar{V}_{\frac{n}{2}2})$. The first match in each pair will be randomized to either treatment or control using the `rbinom` function in R. Next, we subset $\bar{V}_1, \bar{V}_2, \dots, \bar{V}_m$ into appropriate randomization subgroups: $\bar{V}_{1T}, \bar{V}_{1C}, \bar{V}_{2T}, \bar{V}_{2C}, \dots, \bar{V}_{\frac{n}{2}T}, \bar{V}_{\frac{n}{2}C}$ where $\bar{V}_{iT} = (v_{i1}^T, v_{i2}^T, \dots, v_{in}^T)$, similarly for \bar{V}_{iC} . Using these we find

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

THE ABOVE NEEDS A BETTER ABSOLUTE VALUE SYMBOL. for $j = 1, 2, \dots, m$. We randomize N times and find k_{lj} the difference in the two arms for the j^{th} priority variable for each of the $l = 1, 2, \dots, N$ re-randomizations. To assist analysis we draw a parallel coordinates plot where the j^{th} axis plots k_{lj} for $l = 1, 2, \dots, N$. 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 penalty 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 diagnoses, 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.

| Primary import | Secondary import | Tertiary import |
|----------------|------------------|-----------------|
| % DC w/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