1. Introduction
   1. MRTs – Context (why should reader care)
      1. MRTs involve randomization within or between clusters (schools) conducted across multiple geographic sites (schools/districts)
   2. What is the issue with MRTs
      1. Ostensibly increase external validity, however there is ambiguity when understanding what population is being generalized to
   3. Generalizability
      1. Generalizability in the sense of attaining a sample that is representative of a well specified population to inform policy (consider equity)
   4. Current sampling practices
      1. Convenience sampling is typically implemented in MRTs as researchers balance limited resources across all aspects of study design and implementation
   5. Random/probability sampling
      1. The theoretical ideal but is limited by response rates, geographic dispersion, cost of recruitment effort
   6. SBS – potential solution
      1. Stratified balanced sampling is promising and readily accessible method for stepping up your generalizability game, been done a few times
   7. Lack of methodological research in sampling for field trials
      1. SBS Methodologically untested and perhaps overly optimistic
      2. General gaps in literature
   8. Study goal
      1. Need framework for modeling recruitment process to be in the position to study recruitment in field trials
         1. Develop framework for modeling sample selection and school participation
         2. Test SBS and other methods in this framework
2. Results
   1. Generalizability – B-index [plot]
      1. Stratified balanced sampling beats out other methods at participation rates below 50%
      2. Random sampling better than convenience sampling
      3. Stratified random better than simple random
      4. Puzzling features
         1. Stratified convenience sampling is worse than unstratified by this measure
         2. SBS underperforms at higher response rates (degrades)
         3. Motivation to dig into SMDs
   2. Generalizability - ICC vs Coefs [put this piece in methods]
      1. Relationship between ICC and Coef resulting from our simulation specifications
   3. Generalizability – SMD [include values in plots]
      1. Performance seems to be driven by two factors
         1. The extent to which strata capture variation in covariate (ICC)
         2. Strength of relationship between covariate and response rate
      2. For most covariates, stratification improved generalizability [plot]
      3. For two covariates, only SBS resulted in balance [plot]
      4. For two covariates, all methods resulted in balance [plot]
      5. For three covariates, stratified methods performed worse, or SBS performed worse than SCS and SRS
   4. Feasibility – Response rates
      1. SBS sucks convenience sampling rules. [plot]
   5. Feasibility – Gini Plot
      1. Look at cumulative frequency for being selected vs cumulative proportion [plot]

Look at cowplot

Consider knitr