**Protocol for Determining High Quality Randomized Controlled Trial and Quasi-Experimental Design that Would Allow for Causal Inference**

**Taken from WWC Standards Handbook**

**Two-Group Design**

Randomization

--Was there randomization?

Level of Treatment Assignment

--Determine at what level did the treatment assignment occurred?

1. School
2. Teacher
3. Class
4. Student (e.g. lab-based)

Blocking

--Was there blocking before treatment assignment?

--How was blocking done? What covariates were used in blocking?

Attrition[[1]](#footnote-1)

--Was the attrition high?

--If so, was baseline equivalence established for the post-attrition analysis sample?

*Note.*

For cluster design studies, both cluster level attrition and subcluster (e.g. individual within a cluster) should be addressed if analysis was done at the individual (e.g. student level).

Below is the WWC criteria for attrition, where the red zone is unacceptable levels of attrition and green zone is acceptable level of attrition. If attrition exists in the red zone, baseline equivalence should be addressed for the post-attrition analysis sample. WWC is currently updating the attrition criteria.



Baseline Equivalence (for RCT with high student attrition and QED)

--Was there baseline equivalence?

--What covariates examined to establish baseline equivalence?

--If nonequivalence existed, was statistical adjustment (such as regression and ANCOVA) used to addressed the non-equivalence?

*Note.*

These two covariates are the minimum requirements for WWC Postsecondary:

1. Academic Achievement (best measure of prior achievement is taken right before the start of the study, e.g. for postsecondary, they include high school GAP, SAT/ACT scores)
2. Socioeconomic Status (besides family income, other proxies include first-generation college status, free and reduced priced lunch status, parent education level, Pell grant status)

Criteria for Baseline Equivalence

|  |  |
| --- | --- |
| 0 ≤ Effect Size of Covariate ≤ 0.05 | Satisfies baseline equivalence |
| 0.05 < Effect Size of Covariate ≤ 0.25 | Statistical adjustment required to satisfy baseline equivalence |
| Effect Size of Covariate > 0.25 | Does not satisfies baseline equivalence |

For cluster design studies, cohort cluster level measures cannot be used to establish equivalence when analysis is done at the individual level.

Confounding Factors

--Are there confounders in that one study component existed only in one group, such as

1. Teachers in the treatment group are not in the control group?
2. Treatment occurred in a different time period than the control?

Relevant Outcomes

--Does the measure for each relevant outcome satisfy the following:

1. Face validity?
2. Reliability (i.e. internal consistency of at least 0.5 or a test-retest reliability of at least 0.4 or inter-rater reliability of at least 0.5)?
3. Not overly aligned with treatment?
4. Collected in a similar way across both groups?

Analysis

--Was blocking used in the research design? If so, did the analysis addressed it?

*Note.*

In stratified random assignment, WWC accepted methods of adjustment for probability of assignment in randomization are:

1. Estimating a regression model in which the covariate set includes dummy variables that differentiate subsamples with different assignment probabilities
2. Estimating impacts separately for subsamples with different assignment probabilities and averaging the subsample-specific impacts
3. Using inverse probability weights

--Were the students clustered within classroom, teachers, and schools? If so, did analysis address clustering?

*Note.*

The best studies are those that addressed clustering in the analysis. WWC is working on revising the standards for cluster design. WWC corrects for clustering in statistical significance estimates using the method from Hedge (2005). But these studies must be randomized with low cluster and student attrition or baseline equivalence could be established for both randomized studies with student attrition or quasi-experimental studies.

--Were there multiple comparisons?

--If so, how was the correction done for the multiple comparisons?

*Note.*

WWC adopts the Benjamin-Hochberg (BH) correction to account for multiple comparisons.

1. Attrition, as defined by WWC, is when an *outcome variable* is not available for all participants initially assigned to the treatment and comparison groups. If subsampling was done through random selection or selected on characteristics that were clearly determined prior to treatment assignment, then the subsampling is not considered as attrition. [↑](#footnote-ref-1)