

## Questions and Answers About WWC Group Design Standards Version 4.0 for Missing and Imputed Data

The purpose of this document is to provide answers to the questions submitted before and during the What Works Clearinghouse (WWC) technical assistance webinar, “Missing Data in Group Design Studies,” hosted on April 11, 2019. This webinar explained how to handle missing and imputed data in group design studies using WWC Group Design Standards Version 4.0.

This document is meant as a companion to the webinar slide deck and webinar recording. We combined similar questions and rephrased some others for clarity. If additional questions arise, please contact the WWC Help Desk at <https://ies.ed.gov/ncee/wwc/help>.

The end of this document provides links to resources mentioned throughout this document, such as the WWC Standards Handbook 4.0, the WWC Group Design Standards Online Training, and the WWC Online Study Review Guide.

### GENERAL TERMINOLOGY QUESTION

**The term “analytic sample” seems to be used differently throughout the WWC Standards Handbook 4.0. For imputation, the term is only concerned with outcome data, but if the model of analysis includes baseline covariates, then units with missing baseline data will be subject to listwise deletion. In Step 5, we examine baseline equivalence for each analytic sample, which adds additional complexity to understanding use of the term.**

The term “analytic sample” refers to “the subjects from the intervention and comparison groups used to estimate findings,” as noted on page 14 of the WWC Standards Handbook 4.0. One study may report multiple analyses and therefore may potentially include multiple analytic samples. For instance, in the *Math in Focus* study discussed in the webinar, three different analyses and three different analytic samples were reported based on different combinations of missing baseline and outcome data. A reviewer for the WWC would review each analysis separately. One analysis could meet WWC standards while another analysis from the same study could fail to do so. Recognizing that one study may report multiple analytic samples is key in understanding the term.

In addition, it is important to recognize that statistics from different subsamples may be needed to address the missing data review steps. For instance, the webinar used the term “analytic subsample with observed outcome data” to refer to the sample used in computing attrition for Step 2 (i.e., Is the study a low-attrition randomized controlled trial [RCT]?). Hence, the term “analytic sample” still refers to the full sample used to estimate study findings, and the term “subsample” recognizes that only some of the subjects in the analytic sample contribute to the attrition calculations.

Regarding the point about baseline covariates, if the study authors used listwise deletion to address missing baseline data, then the analytic sample would not have subjects with missing baseline data because those subjects were removed from analyses by definition. However, the study authors could choose to impute missing baseline data to estimate study findings, which means the analytic sample would include subjects without observed baseline data in that case.

## QUESTIONS ABOUT STEP 1: ACCEPTABLE METHODS FOR ADDRESSING MISSING DATA

**Does the table of acceptable approaches (from discussion of Step 1) exist somewhere online as a resource I can share with colleagues?**

We have included this table below and you can also find a PDF on the WWC website.

### Acceptable Approaches for Addressing Missing Data Under WWC Group Design Standards 4.0

Method	Study design	Can be used to include participants with missing...		Additional Requirements (see below)
		Baseline data	Outcome data	
Complete case analysis	All			
Regression imputation	All	✓	✓	3 requirements for imputation model
Dummy imputation	Non-compromised RCTs only*	✓		
Maximum likelihood	All	✓	✓	Use standard statistical package or include relevant citations
Non-response weights	All		✓ <sup>†</sup>	2 requirements for missing outcome data models

**Table notes**

\*However, for QEDs and compromised RCTs, dummy imputation can still be applied to baseline measures NOT specified in the review protocol as required to assess baseline equivalence. A compromised RCT occurs when different analytic choices such as changing a subject's group membership after random assignment compromise the validity of an otherwise well-executed random assignment process.

<sup>†</sup>With non-response weights, participants without observed outcome data will not be included in impact estimation models, but participants with observed outcome data will be weighted so that they resemble the full sample with and without outcome data.

**Additional requirements**

Regression imputation: The imputation regression model must (a) be conducted separately by condition or include an indicator variable for condition, (b) include all covariates used for adjustment in the impact model, and (c) include the outcome when imputing missing baseline data.

Non-response weights: The missing outcome data model must (a) estimate probabilities of missingness separately by condition or include an indicator variable for condition and (b) include all baseline measures specified in the review protocol as required for baseline equivalence.

**Source**

WWC Standards Handbook Version 4.0, pages 39-40. [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_standards\\_handbook\\_v4.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf)

**If we have complete observations that are powerful enough to conduct appropriate analyses, there is no need for imputation and that does not cause any nonresponse bias, correct?**

Both complete case analysis and imputation-based approaches are acceptable methods for addressing missing data under WWC Group Design Standards Version 4.0.

Page 38 of the WWC Standards Handbook 4.0 provides additional description for why complete case analysis is considered acceptable: “All but one of the acceptable approaches in Table II.6 can provide unbiased estimates of the effectiveness of an intervention based on the assumption that the missing data do not depend on unmeasured factors. The exception is complete case analysis, which requires a more restrictive assumption that the missing data also do not depend on measured factors. Because of this, many researchers have recommended against using complete case analysis to address missing data [...]. Nevertheless, the WWC considers complete case analysis to be an acceptable approach for addressing missing data, because possible bias due to measured factors can be assessed through the attrition standard and WWC’s baseline equivalence requirement.”

Hence, the WWC standards recognize that complete case analysis can introduce bias, but that bias can be diagnosed through the attrition standard and baseline equivalence requirement. The standards describe a range of acceptable methods, and researchers can use their discretion to pick which acceptable method that they think is best suited to a particular study context.

### **What defines a "noncompromised RCT"?**

A compromised RCT occurs when various analytic choices and study team decisions compromise the validity of an otherwise well-executed random assignment process. The WWC considers the random assignment process to be compromised under four different scenarios:

1. The study includes subjects in the sample used to estimate findings (analytic sample) who were not randomly assigned.
2. Subjects are randomly assigned to a group with different probabilities, but the findings are based on an analysis that does not account for the different assignment probabilities.
3. A study author changes a subject’s group membership after random assignment.
4. A study author manipulates the analytic sample to exclude certain subjects based on events that occurred after the introduction of the intervention when there is a clear link between group status and the reason for the exclusion.

Pages 7–9 in the WWC Standards Handbook 4.0 and Module 1 (Group Designs) in the online training describe these four scenarios in more detail.

**For the *Math in Focus* study discussed during the webinar, why did the researchers not impute all the missing data? Can you impute data if a person has unit missing data at both baseline and outcome?**

In the *Math in Focus* study discussed during the webinar, the study authors choose not impute all missing data using multiple imputation because “a whole randomized block (school) dropped out of the study just after randomization, yielding neither outcomes nor covariate data for that school” (Jaciw et al., 2016, p. 483).

The WWC’s missing data standards do not explicitly prohibit imputing data for subjects with missing baseline and missing outcome data. But the uncertainty in imputing data for those subjects will be higher than for subjects with at least some observed baseline or outcome data. Hence, some study authors may choose to exclude subjects with missing baseline and outcome data from analysis but impute data for other subjects. This approach would be considered acceptable under WWC Group Design Standards Version 4.0, assuming other relevant requirements are satisfied such as the imputation regression model must include all covariates used for statistical adjustment in the impact estimation model (see pages 39–40 in the WWC Standards Handbook 4.0).

**How acceptable is multiple imputation in various social science fields? Do journals now expect researchers to use multiple imputation?**

From the WWC’s perspective, under WWC Group Design Standards Version 3.0, only low-attrition RCTs could use multiple imputation and meet the standards. However, now under Version 4.0, multiple imputation can be used in all eligible group design studies, which is consistent with methodological research indicating the method's appropriateness in a wide range of scenarios.

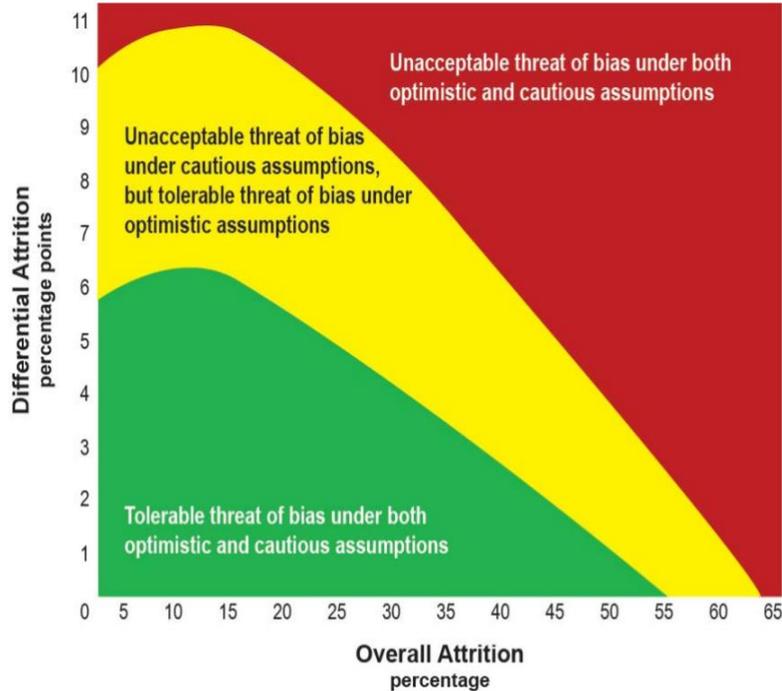
**QUESTIONS ABOUT STEP 2: IS THE STUDY A LOW-ATTRITION RCT?**

**How is a “low-attrition RCT” defined? Is there a fine cut-off for that?**

The procedures for assessing attrition can be found in pages 9-14 of the WWC Standards Handbook 4.0 and Module 2 (Attrition) of the WWC Group Design Standards Online Training (<https://ies.ed.gov/ncee/wwc/onlinetraining>).

As described in those resources, the definition of a low-attrition RCT depends on three main considerations: (a) overall attrition rate, (b) differential attrition rate across experimental conditions, and (c) the review protocol’s specification of whether optimistic or cautious assumptions should be used about the potential threat of bias from attrition. Figure II.2 in the WWC Standards Handbook 4.0 summarizes these three considerations.

Figure II.2. Attrition and Potential Bias



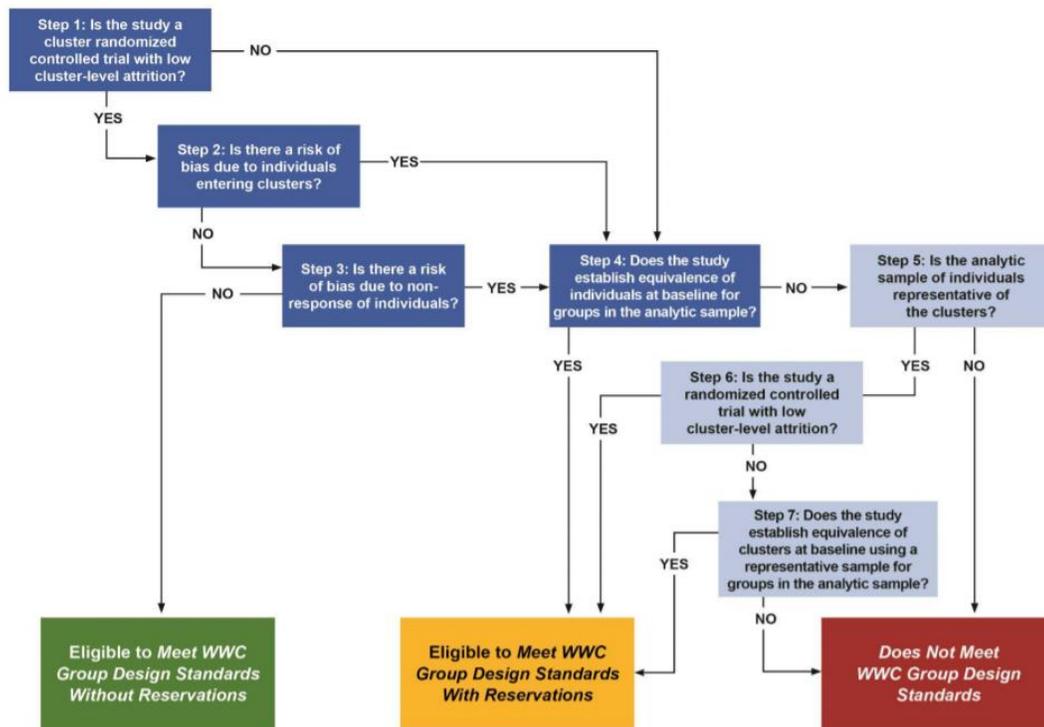
Note: Not every combination of differential and overall attrition is possible for any given study. The review protocol will specify which set of attrition boundary values applies.

**How is a “low-attrition RCT” defined for studies using cluster-level assignment? For cluster RCTs, is attrition based on the unit of random assignment (e.g., schools) or the unit of the outcome measure (e.g., students)?**

Although the WWC’s missing data standards as described on pages 36–46 of the WWC Standards Handbook 4.0 currently do not provide explicit guidance about defining a low-attrition cluster RCT, other portions of the WWC Standards Handbook 4.0 are relevant. Specifically, pages 22–27 and Figure II.4 (reproduced below) note that three conditions must be met for a cluster RCT to receive the highest study rating, *Meets WWC Group Design Standards Without Reservations*:

- Step 1: Is the study a cluster randomized controlled trial with low cluster-level attrition? The answer must be yes for the study to receive the highest rating.
- Step 2: Is there a risk of bias due to individuals entering clusters? The answer must be no for the study to receive the highest rating.
- Step 3: Is there a risk of bias due to nonresponse of individuals? The answer must be no for the study to receive the highest rating.

Figure II.4. Review Process for Cluster-Level Assignment Studies



If any of those three conditions are not met, then the highest rating a cluster RCT can receive is *Meets WWC Group Design Standards With Reservations*. For instance, even if cluster-level attrition is low, the study cannot meet WWC standards without reservations if it has high individual-level nonresponse. Page 26 of the WWC Standards Handbook 4.0 defines the term *individual-level nonresponse* and explains why that term is preferred over *individual-level attrition* in the context of cluster RCTs. Module 8 (Cluster-Level Assignment) of the WWC online training also covers these issues in more detail (<https://ies.ed.gov/ncee/wwc/onlinetraining>).

The WWC’s *Standards and Procedures Handbook* advisory team plans to reexamine the missing data standards for cluster RCTs to provide more explicit guidance and make iterative improvements in the coming years. In the meantime, a conservative approach is to assume that the three conditions mentioned above must be met to answer “yes” to the “Is the study a low-attrition RCT?” missing data review step for cluster RCTs.

**Even if we impute data, do missing data still count as attrition when computing attrition?**

When calculating attrition rates, imputed outcome data count as attrition because both imputed and missing outcome data represent a potential threat of bias, as noted on page 41 of the WWC Standards Handbook 4.0. However, imputed baseline data do not count as attrition because randomization helps ensure baseline equivalence in the full randomized sample.

Knowledge Check 2 in the webinar addressed this issue by considering a scenario involving both imputed baseline and imputed outcome data. In this example, the analytic sample included

subjects with observed baseline or observed outcome data. But the sample used to compute attrition rates would be the analytic subsample with observed outcome data because imputed outcome data, but not imputed baseline data, count as attrition.

### **QUESTIONS ABOUT STEP 3: DOES THE STUDY LIMIT POTENTIAL BIAS FROM IMPUTED OUTCOME DATA?**

#### **Where is the information for how to assess bias from imputed outcome data?**

Pages 41–43 and Appendix B of the WWC Standards Handbook 4.0 provide details on implementing Step 3, “Does the study limit potential bias from imputed outcome data, if any outcome data are imputed?” In addition, the online SRG will automatically apply this step after the relevant information (e.g., outcome means for different samples) is entered.

#### **When estimating potential bias from imputed outcome data, should the baseline means be calculated within the analytic framework (e.g., HLM) or simply be regular means?**

In general, the means to complete the missing data review steps should be unadjusted means, not covariate-adjusted means. For instance, the question above references assessing potential bias from imputed outcome data, which requires knowing the baseline means for different samples, as noted on page 42 of the WWC Standards Handbook 4.0. These baseline means should be unadjusted means, not controlling for the outcome or other baseline variables.

The question also asks how to compute these means if hierarchical linear models (HLMs) will be used to estimate study findings. Although the missing data standards as described in pages 36–46 of the WWC Standards Handbook 4.0 do not explicitly address this issue, we can look to other relevant portions of the Handbook for guidance:

- Page 27: “Regardless of the level of analysis, this baseline equivalence requirement must be satisfied using individual-level standard deviations. Means calculated using either cluster- or individual-level data are acceptable as long as the weighting is consistent with the weighting used in the analysis.”
- Page 17: “If the study used weights in the analysis, the baseline means must also be calculated using the same weights.”

These sentences are relevant because applying HLM involves reweighting based on the clustered nature of the data. Hence, using means based on weights from HLMs would be consistent with this guidance from other sections of the WWC Standards Handbook 4.0, even though the section on missing data does not explicitly require weighted means for this scenario. The WWC’s *Standards and Procedures Handbook* advisory team plans to examine the issue and provide more concrete guidance in the coming years.

**QUESTIONS ABOUT STEP 5: (A) DOES THE STUDY SATISFY BASELINE EQUIVALENCE FOR THE ANALYTIC SAMPLE? (B) IS BASELINE EQUIVALENCE SATISFIED ACCOUNTING FOR MISSING OR IMPUTED BASELINE DATA?**

**In the scenario where analysts used/reported nonimputed baseline measures to assess baseline equivalence, but used imputed versions of those measures as covariates in the analysis, would it be correct or fine for the reviewer to use Step 5a for assessing baseline equivalence?**

The answer is no, assuming the baseline measure is specified in the review protocol as required for assessing baseline equivalence. Step 5b, not 5a, would need to be applied in that case. Baseline equivalence must be demonstrated for the analytic sample, which in this case included participants with imputed baseline data, even if the analysts reported baseline differences based on observed baseline data. Step 5b provides explicit guidance for this scenario.

Two samples should be distinguished from each other:

- Analytic sample: subjects used to estimate study findings. This sample included subjects with imputed baseline data in this case.
- Complete case subsample: subjects with observed baseline and outcome data. This sample is a subsample of the analytic sample. Use of the term *complete case* here does not imply that complete case analysis was used to estimate effects on outcomes.

In describing Step 5b, Appendix C in the WWC Standards Handbook 4.0 uses the term *complete case baseline effect size* ( $g_{xR}$ ) for the baseline effect size mentioned in the preceding question. This type of baseline effect size is distinct from the baseline effect size based on imputed data ( $g_{xI}$ ).

The review would proceed as follows:

- Step 5. Are data in the analytic sample missing or imputed for any baseline measure specified in the review protocol?
  - Yes, the sample used to estimate findings included participants with imputed baseline data. Therefore, apply Step 5b.
- Step 5b. Does the study satisfy baseline equivalence using the largest baseline difference accounting for missing or imputed baseline data?
  - Apply Equations C1–C4 and C1\*–C4\* in Appendix C of the WWC Standards Handbook 4.0 because the estimate of the baseline difference is based on the complete case subsample.
  - If the estimate of the baseline difference was instead based on imputed baseline data, Equations D1–D4 and D1\*–D4\* in Appendix C of the WWC Standards Handbook 4.0 would be applied.

Step 5a would only be applied if the authors restricted their analysis to participants who had observed data for baseline measures specified in the review protocol as required for assessing baseline equivalence. For instance, Step 5a would be applied if complete case analysis was used to estimate effects on outcomes.

## MISCELLANEOUS QUESTIONS

**Suppose the outcome measure is a survey-based scale constructed from multiple survey items, some (but not all) of which are missing for some participants. Does what we have learned apply when we impute some of the missing individual survey items, as opposed to imputing the entire measure?**

This important question requires more consideration. Currently, the WWC Group Design Standards Handbook 4.0 does not provide explicit guidance regarding this issue. We anticipate this issue will be addressed in future versions of the Handbook.

**If the *Math in Focus* study discussed during the webinar included the outcome in the multiple imputation model, what would be the rating?**

If the *Math in Focus* study discussed during the webinar had used an acceptable imputation model in the multiple imputation analyses (i.e., the imputation model also included the outcome when imputing baseline data), then two steps would need to be assessed:

- Step 3: Does the study limit potential bias from imputed outcome data?
- Step 5b: Does the study satisfy baseline equivalence using the largest baseline difference accounting for missing or imputed baseline data?

If the answer is yes to both questions, then the study's multiple imputation analyses would be eligible for the *Meets WWC Group Design With Reservations* rating because the study is a high-attrition RCT. If the answer is no to either Step 3 or Step 5b, then the multiple imputation analyses would be rated *Does Not Meet WWC Group Design Standards*.

**What software packages can be used to implement the methods and review procedures discussed in the missing data standards?**

The missing data methods discussed in the WWC Standards Handbook 4.0 such as regression imputation can be implemented in most standard statistical software packages. Researchers should feel free to use any software package that they feel most comfortable with. Then, to diagnose potential bias, they can enter the relevant information such as means and standard deviations in the online SRG (<https://ies.ed.gov/ncee/wwc/Studyreviewguide>), which will automatically apply the WWC's review procedures.

Alternatively, advanced users could take the formulas in the WWC Standards Handbook 4.0's Appendices B and C and use them in an environment in R. However, this process requires

advanced programming skills and careful examination of mathematical formulas. The online SRG provides an alternative for users who do not wish to do that.

**If I use a large-scale secondary data, do I have to deal with missing data?**

The answer is yes if there is missing or imputed data, which is likely. Analyses based on primary data versus large-scale secondary data are reviewed the same. However, one special consideration for analysts is that large-scale secondary datasets may be released with imputation procedures already applied. Hence, the analyst should carefully review the data preparation manuals to ensure adherence to WWC standards.

**Is there a threshold for determining if there is too much missing data?**

The WWC Group Design Standards Version 4.0 do not have a set maximum threshold for an acceptable percentage of missing data. However, other thresholds and boundaries are relevant:

- Step 2: Is the study a low-attrition RCT? The boundaries for defining high versus low attrition are given in pages 9–14 of the WWC Standards Handbook 4.0. However, high-attrition RCTs can still meet WWC standards with reservations if later steps are satisfied.
- Step 3: Does the study limit potential bias from imputed outcome data? The potential bias due to imputed outcome data must be less than 0.05 standard deviations.
- Step 5: Does the study satisfy baseline equivalence? The largest estimated baseline difference must not exceed 0.25 standard deviations if the analysis includes an acceptable adjustment for the baseline measure or 0.05 standard deviations otherwise.

## GENERAL RESOURCES

In addition to the webinar held on April 11, 2019, the following resources provide guidance for reviewing studies with missing or imputed data under the WWC Group Design Standards Version 4.0.

- **WWC Standards Handbook 4.0**
  - [https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc\\_standards\\_handbook\\_v4.pdf](https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf)
  - Pages 36–46, Appendix B, and Appendix C detail standards and procedures for reviewing analyses with missing or imputed data
- **WWC Group Design Standards Online Training**
  - <https://ies.ed.gov/ncee/wwc/onlinetraining>
  - Module 1 (Group Designs), Module 2 (Attrition), and Module 3 (Baseline Equivalence) provide more general introductions to topics that are especially important to understanding the WWC’s missing data standards.
- **WWC Online Study Review Guide (SRG)**
  - General landing page: <https://ies.ed.gov/ncee/wwc/Studyreviewguide>
  - Public online SRG: <https://ies.ed.gov/ncee/wwc/wwcsrgpublic>
  - After you enter relevant information about a study, the online SRG will automatically apply the steps of the missing data review process and tell you the WWC study rating at the end.
- **WWC Recertification for Standards Version 4.0 Prior Webinar**
  - <https://www.youtube.com/watch?v=LkU42QtnpMc>
  - Prior webinar focused more broadly on changes from Version 3.0 to Version 4.0 WWC Group Design Standards.
  - This webinar’s content on missing and imputed data (e.g., knowledge checks and figures) formed much of the basis for this new webinar.

## REFERENCE

Jaciw, A. P., Hegseth, W. M., Lin, L., Toby, M., Newman, D., Ma, B., & Zacamy, J. (2016). Assessing impacts of *Math in Focus*, a “Singapore Math” program. *Journal of Research on Educational Effectiveness*, 9(4), 473–502. <https://doi.org/10.1080/19345747.2016.1164777>