IES Institute of Education Sciences

Questions and Answers About Effect Size and Standard Error Formulas

The purpose of this document is to provide answers to questions submitted before and during the What Works Clearinghouse (WWC) technical assistance webinar "Effect Size and Standard Error Formulas." The webinar was hosted on December 2, 2020.

This document is meant to serve as a companion to the webinar slide deck and webinar recording. The WWC webinar team combined similar questions and rephrased others for clarity while preserving the meaning of the original questions. If additional questions arise, please contact the WWC Help Desk at https://ies.ed.gov/ncee/wwc/help.

1. **Can the regression-adjusted standard error approach accommodate multiple imputation?**

   The regression-adjusted standard error approach can accommodate multiple imputation, provided that the authors' reported coefficient standard error properly adjusts for the uncertainty from the imputation process. As noted in table II.6 in version 4.1 of the *WWC Standards Handbook*, the calculation of the reported coefficient standard error must meet two requirements:

   a. It must be based on at least five sets of imputations, and

   b. It must account for both the within-imputation and between-imputation variance components, as well as the number of imputations.

   Most established multiple imputation routines satisfy these requirements.

   If the analysis imputed outcome data for a cluster-level assignment study, the study must also provide evidence that the approach appropriately adjusts standard errors for clustering by citing a peer-reviewed journal article or textbook that describes the procedures and demonstrates their effectiveness in accounting for clustering.

2. **How would you approach a situation in which the authors of a study reported an effect size, but it could not be replicated using the information provided for the study? For example, what strategy would you use if the WWC calculated an effect size based on the information provided for the study, but it differed from the author-reported effect size?**

   The WWC generally reports the WWC-calculated effect size because its computation can be verified, and it supports comparability across outcomes and studies. However, in some cases, the WWC will use the study-reported effect size if **all** the following conditions are true:

   a. The study calculated the effect size consistent with a WWC formula for Hedges' *g*, including a small sample adjustment for continuous outcomes.

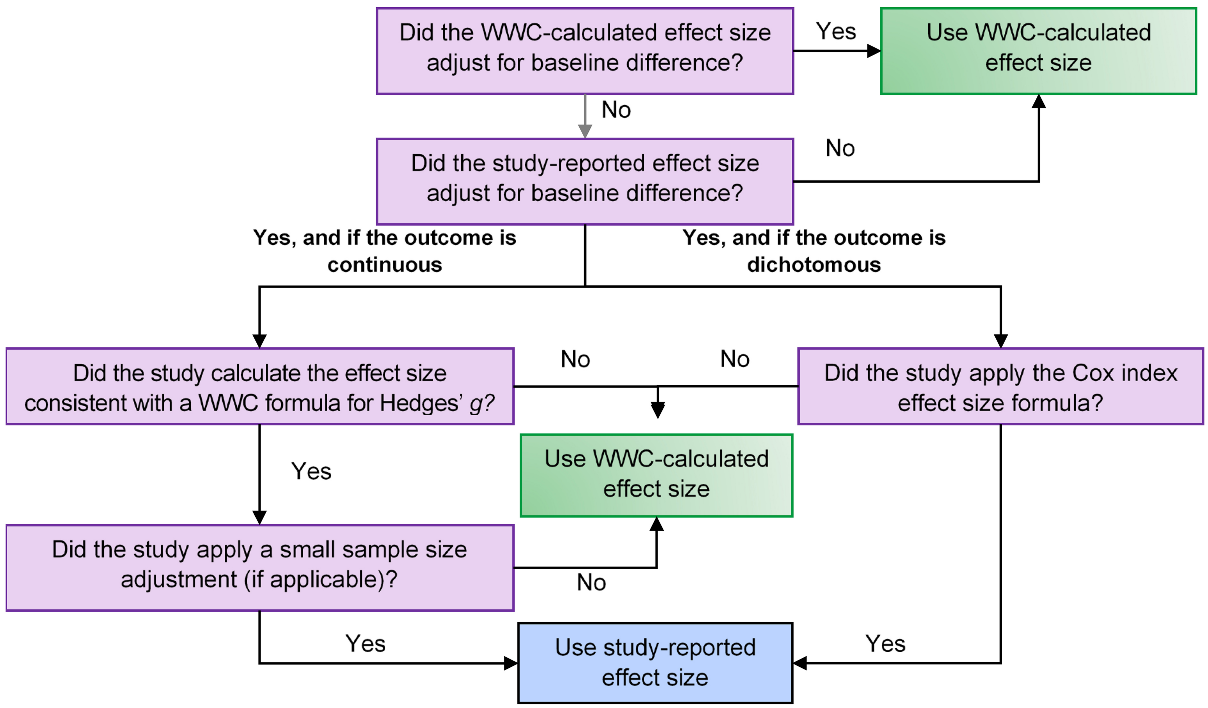
   b. The study-reported effect size regression adjusts for baseline differences, and the WWC-calculated effect size does not adjust for baseline differences or the WWC-calculated effect size is based on a post hoc difference-in-differences adjustment.

c. The study used the Cox index for dichotomous outcomes, as appropriate.

Whether the WWC uses the author-reported effect size or the WWC-calculated effect size, the WWC will report this information in the Data From Individual Studies database: https://ies.ed.gov/ncee/wwc/StudyFindings.

The following figure summarizes this logic:

Figure 1. When will WWC use the study-reported effect size?



3. **Will the WWC use the new methods to override study-reported *p*-values?**

The principles for selecting WWC-calculated versus study-reported *p*-values in version 4.1 of the *WWC Standards Handbook* remain the same as in version 4.0. These principles state that, in contrast to study-reported effect sizes, the WWC generally accepts the study-reported *p*-values and statistical significance (see page 9 of the *WWC Procedures Handbook, Version 4.1*).

However, there are two common circumstances in which the WWC will favor the WWC-calculated *p*-value over the study-reported one:

a. The study does not include statistical significance estimates, or there is a known problem with the study calculations; or

b. The statistical significance levels reported in the study do not account for clustering when there is a mismatch between the unit of assignment and the unit of analysis.

4. **Why does the WWC use a fixed effects model rather than a random effects model for the synthesis of findings from different studies?**

   Implementing a random effects model requires estimating the between-studies variance component, and that estimate can be unreliable when the meta-analysis includes only a few studies. Most WWC intervention report syntheses are based on a few studies. The fixed effects (plural) model, on the other hand, enables the WWC to conceptually acknowledge variability in effects. That is, the observed effects, and their variability, are representative of a larger population of effects. Furthermore, fixed effects meta-analysis offers a more robust approach to synthesis than the previous vote-counting approach.

5. **I am a study author concerned with getting the highest WWC rating. Do I need to use the new formulas to achieve the highest rating?**

   Study authors do not have to use the statistical formulas from version 4.1 of the *WWC Procedures Handbook* to achieve the highest rating. Authors should, however, report all the information necessary to apply these formulas. For example, if the authors used multiple regression to estimate an intervention effect, they should, at a minimum, report the following:

   a. The unstandardized coefficient for the intervention effect estimate;

   b. The coefficient standard error; and

   c. The raw, unadjusted standard deviations separately for the intervention and comparison groups.

   The coefficient standard error is not required to compute WWC effect sizes or standard errors, but reporting it helps ensure that the WWC's calculations align with the authors' calculations.

   Providing effect sizes in study reports should be considered good statistical practice, but the WWC will verify these effect size computations using other reported information regardless.

6. **In cluster-level assignment studies, does the total number of individuals or total number of clusters determine whether the small-sample corrections should be applied?**

   The WWC applies small sample size adjustments to its effect size calculations regardless of sample size. For large numbers of individuals and large numbers of clusters, these adjustments will be minor, but the WWC will still apply them.

7. **What values will the WWC use for the intraclass correlation coefficient?**

   The WWC will use study-reported intraclass correlation coefficient values when available. If the study-reported value is unavailable, the WWC will generally use 0.20 for achievement outcomes and 0.10 for all other outcomes, unless otherwise specified in the review protocol (see page 20 of the *WWC Procedures Handbook, Version 4.1*). These default values are based on empirical literature in the field of education research (Hedges & Hedberg, 2007; Schochet, 2008).

8. **Can you confirm that the small-sample correction term ω shows up twice in the standard error formula on slide 24, once explicitly and once within the g^2 on the second term under the radical?**

   The standard error formula on slide 24 does include the small-sample correction term ω twice, both explicitly on the left side and then implicitly in the computation of the right-hand g^2 term.

9. **I am concerned that covariate-adjusted effect sizes might not be comparable to the standard effect sizes. In a case in which the mean values of the covariates are notably different between intervention and control groups, these two could diverge substantially. Similarly, studies using different lists of covariates could produce different effect sizes even if the actual effect were the same in the studies. This is why traditional meta-analyses don't combine the partial effect sizes unless you are sure the same covariates are used in all the studies.**

   Effect sizes from intervention studies that are adjusted for pre-intervention covariates help mitigate observable sources of selection bias and improve the precision of the estimates. That is, covariate adjustment in the types of studies the WWC reviews helps model the selection process. In nonintervention studies (such as studies of bivariate or multivariate relationships), covariate adjustment that affects the magnitude of the effect size indeed has implications for comparability across varying adjustment strategies, especially because the breadth of potential confounding factors is much greater.

   The WWC *requires* quasi-experimental designs (QEDs) and high-attrition randomized controlled trials (RCTs) to adjust outcomes for specific pre-intervention covariates if baseline differences on those covariates are 0.05 standard deviations or greater. Covariates used in QEDs and high-attrition RCTs may change the magnitude of the effect size, to the degree that selection into the intervention is related to observable covariates and when those covariates are related to outcomes. However, the required adjustment strategies for QEDs operate as a minimum set of standards for mitigating selection bias due to observable factors, and researchers may identify and measure additional selection threats that warrant further adjustment.

   Conversely, the WWC prefers—but does not require—low-attrition RCTs to adjust for pre-intervention covariates. Covariates used in RCTs are not expected to systematically affect the magnitude of the effect size, as receiving the intervention should be unrelated to observable and unobservable factors at baseline. But controlling for covariates can still increase statistical precision of intervention effect estimates in low-attrition RCTs.

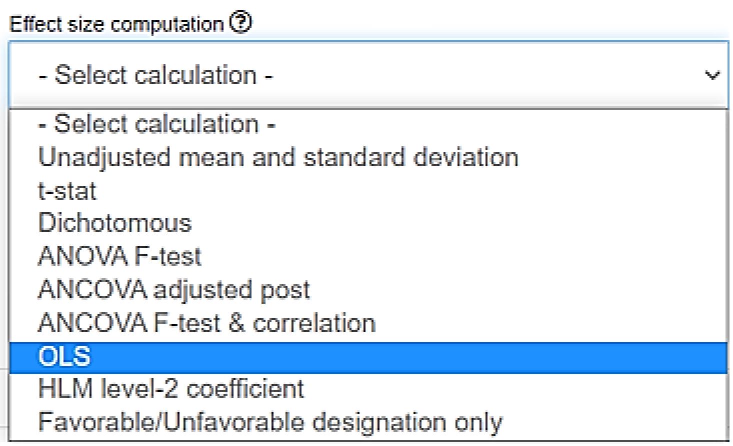
   Therefore, by using the covariate-adjusted effect size, the WWC is keeping with the historical precedent.

10. Do you need to be an official WWC reviewer to access the Online Study Review Guide (OSRG)?

Members of the public can select WWC Group Design Standards versions 3.0 and 4.0 in the public OSRG at https://ies.ed.gov/ncee/wwc/StudyReviewGuide. This public OSRG does not yet implement version 4.1. As of December 23, 2020, official WWC reviewers can conduct version 4.1 reviews using the restricted-access OSRG at https://members.nces.ed.gov/OSRG.

11. Will version 4.1 of the OSRG include notes on the Rating page that describe which effect size standard error calculation was used? This would help reconcilers confirm that reviewers coded the outcome measures correctly.

In the OSRG, reviewers will choose an option in the *Effect size computation* dropdown menu that determines which pieces of statistical information the OSRG will use to compute WWC effect sizes and standard errors (see screenshot below).



This choice will be reflected as an endnote in the outcome details tables in the OSRG, including for version 4.1. The screenshot below shows an example for the ordinary least squares (OLS) regression coefficient option (note at the bottom of the screenshot below).

	Official	Study-reported	Calculated
Effect size	0.318 *		0.318 *
Standard error	0.139 *		0.139
Improvement index	+12		
p value	0.023		0.023
Significant	Yes		Yes

* The effect size was calculated using the OLS calculation (coefficient=7.89).

In many cases, the effect size computation option also directly corresponds to the standard error formula that the OSRG will use. In some cases, however, the OSRG's application of standard error formulas may depend on other entered information as well.

Consider the OLS option as an example:

- If either the coefficient standard error or coefficient test statistic is entered, then the OSRG will use the regression-adjusted standard error approach noted in the supplement to the *WWC Procedures Handbook, Version 4.1* (see equations E.7.0 and E.7.1).
- If both the coefficient standard error and coefficient test statistic fields are left blank, then the OSRG will use the effect size standard error formula based on the proportion variance explained (R^2); see equation E.2.2 in the *WWC Procedures Handbook, Version 4.1*. (The OSRG will assume $R^2 = 0$ if the R^2 field is also left blank.)
- If both the regression coefficient standard error (or coefficient test statistic) and R^2 value are entered, the OSRG will use the regression-adjusted standard error approach.

The current OSRG functionality adds a table note about the effect size computation option used (such as the OLS option), but as of December 2020, this table note functionality does not note these finer-grained standard error distinctions. Reviewers may consult appendix E of the *WWC Procedures Handbook, Version 4.1*, and its supplement to better understand the underlying logic the OSRG uses to compute standard errors.

12. Does the WWC have standard error formulas that account for covariates in analyses of dichotomous outcomes?

An ongoing area of development for the WWC is calculating standard errors for covariate-adjusted analyses of dichotomous outcomes. Various complexities that need to be considered include how covariate adjustment may change the interpretation of logistic regression coefficients (such as differences between marginal and conditional odds ratios), how clustering assumptions operate for dichotomous outcomes in cluster-level assignment studies, and how to best address models that ignore the boundedness of the outcome (such as dichotomous outcomes analyzed with OLS models).

In the meantime, the WWC's current standard error formulas for dichotomous outcomes assume simple, unadjusted proportions (see page E-7 of the *WWC Procedures Handbook, Version 4.1*).

13. Can you further explain the cluster-level assignment extension for the standard error formula based on R^2 values? I would assume that the R^2 value should correspond to the level-2 R^2 (that is, proportion of level-2 variance explained by the covariates).

The supplement to the *WWC Procedures Handbook, Version 4.1*, includes the following cluster-level assignment extension to the standard error formula based on R^2 values (see equation E.2.2 in table 3 of the supplement):

$$\omega\sqrt{\frac{n_i + n_c}{n_i n_c}(1 - R^2)\eta + \frac{g^2}{2h}}$$

where ω is a small-sample correction term, n_i and n_c are the total number of individuals in the intervention and comparison groups, η is the cluster design effect, and h is the degrees of freedom.

This formula is intended for single-level models such as OLS models that ignore clustering or OLS models that use cluster-robust standard errors. Version 4.1 of the *WWC Procedures Handbook* currently does not include a standard error formula that uses the level-1 and level-2 R^2 values from more complex, variance-partitioning models such as hierarchical linear models.

For future handbook versions, the WWC's Statistical, Technical, and Analysis team will consider further extensions for cluster-level assignment studies that could factor in other statistical information (such as level-1 and level-2 R^2 values).

Related Resources

In addition to the webinar and this Questions and Answers document, the following resources provide guidance about effect size and standard error formulas, as well as the WWC's standards, procedures, and review protocols. These include resources shared in the chat box during the webinar.

- *Supplement to the WWC Procedures Handbook, Version 4.1:* https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-41-Supplement-508_09212020.pdf
- *WWC Standards Handbook, Version 4.1*: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf
- *WWC Procedures Handbook, Version 4.1*: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf
- *WWC database of data from individual studies:* https://ies.ed.gov/ncee/wwc/StudyFindings
- Sign up to receive Institute of Education Sciences news briefs: https://ies.ed.gov/newsflash/
- *WWC Help Desk:* https://ies.ed.gov/ncee/wwc/Help

References

Schochet, P. Z. (2008). *Statistical power for random assignment evaluations of education programs.* Journal of Educational and Behavioral Statistics, 33(1), 62–87. https://eric.ed.gov/?id=EJ788461

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://eric.ed.gov/?id=EJ782420