IES Institute of Education Sciences

# Supplement to the What Works Clearinghouse Procedures Handbook, Version 4.1

What Works Clearinghouse™

NCEE
U.S. DEPARTMENT OF EDUCATION

*A Publication of the National Center for Education Evaluation at IES*

## A. Overview

This supplement concerns Appendix E of the *What Works Clearinghouse* (WWC) *Procedures Handbook, Version 4.1*. The supplement extends the range of designs and analyses that can generate effect size and standard error estimates for the WWC. In Version 4.1, the WWC added standard error formulas to accommodate the fixed-effects meta-analysis approach for combining and weighting effect size estimates across studies (see Appendix H in the *WWC Procedures Handbook*). The WWC also uses these standard errors when computing the WWC-calculated $p$-values for individual effects (see Section VI.A.2).

This supplement presents several new standard error formulas for cluster-level assignment studies, in addition to introducing a needed bias correction term to the cluster design effect size formula. However, some extensions also expand options for computing standard errors for individual-level assignment studies. This supplement introduces a general approach, for both individual- and cluster-level assignment studies, that leverages the study-reported standard errors for regression coefficients.

These enhancements were identified after the initial publication of the *WWC Procedures Handbook, Version 4.1,* in January 2020. No reviews were conducted under Version 4.1 prior to the publication of this supplement. This supplement is specific to Version 4.1, and its content will be directly integrated into future versions of the *WWC Procedures Handbook*.

## B. Defining effect sizes for studies with cluster-level assignment

Appendix E of the *WWC Procedures Handbook, Version 4.1,* shows that the effect size of interest for cluster-assignment designs uses the total variability, both between- and within-clusters, as the standardizer (see equation E.5.2). However, the current equation E.5.1 for the effect size formula results in slightly upwards-biased estimates of the population effect size of interest (see Hedges, 2007, equations 3 and 15[1]), when the intraclass correlation (ICC)[2] is greater than zero. This supplement therefore amends equation E.5.1 with another bias-correction term, as shown in Table 1.

**Table 1. Amendment to the effect size formula for cluster-assignment designs**

| Eq. number | Version | Formula |
|---|---|---|
| E.5.1 | Biased formula | $g = \frac{\omega b}{S}$, where $S = \sqrt{\frac{(n_i-1)s_i^2+(n_c-1)s_c^2}{n_i+n_c-2}}$ |
| | Unbiased formula | $g = \frac{\omega b}{S}\sqrt{1-\frac{2(n-1)\rho_{ICC}}{N-2}}$ |

*Note*. $g$ is the standardized effect size estimate; $b$ is an estimate of the unstandardized difference between intervention and comparison group means; $S$ is the pooled standard deviation defined in equation E.1.1 in Appendix E; $N = n_i + n_c$, or the total sample of individuals; $n$ is the average number of individuals per cluster; $\rho_{ICC}$ is the intraclass correlation coefficient.

[1] Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://eric.ed.gov/?id=EJ782531

[2] This supplement uses $\rho_{ICC}$ notation for the ICC to distinguish it from another correlation that is referenced in this document, $\rho_{cor}$, which is the correlation between baseline and outcome measures. The Handbooks uses a general $\rho$ notation for both of these values.

The added term can be viewed as a small number of clusters adjustment. It slightly reduces the effect size estimate, usually by less than 1% in most WWC review contexts. Although typically small, this adjustment is necessary to ensure alignment with the standard error formulas and to obtain valid WWC-calculated $p$-values. This small number of clusters correction is in addition to the small sample size adjustment $\omega$, which is equal to $1 + 3/(4df - 1)$, where $df$ is the degrees of freedom for cluster-level assignment studies (see equation F.1.2 in Appendix F). The WWC will follow existing guidelines for imputing ICC values when they are not available from the original study reports (see Appendix F for more detail).

## C. Regression-adjusted standard errors for cluster-level assignment studies

Intervention studies commonly use multiple regression models to adjust for baseline covariates and to increase the precision of intervention effect estimates. To address the role of such covariates, this supplement introduces a new approach for calculating effect size standard errors. It leverages study-reported information about the unstandardized regression coefficient for the adjusted group mean difference (represented as $b$ in Table 1).

For cluster designs, if the authors' regression model properly adjusted for clustering, the standard error for the standardized effect size estimate $g$ is given by

[E.7.0] $$SE[g] = \omega\sqrt{\left(\frac{SE_{CC}}{S}\right)^2 \gamma + \frac{g^2}{2h}}$$

where $SE_{CC}$ is the cluster-corrected standard error for the unstandardized regression coefficient, $S$ is the pooled standard deviation, $\gamma = 1 - \frac{2(n-1)\rho_{ICC}}{N-2}$ is the small number of clusters correction, $g$ is the effect size estimate, and $h$ is the degrees of freedom. The original study authors can adjust for clustering in multiple acceptable ways, including using hierarchical linear modeling (HLM) or applying cluster-robust standard errors to ordinary least squares (OLS) models (for more detail, see Appendix F).

## D. Extending the regression-adjusted standard error approach

The regression-adjusted standard error approach can be extended to accommodate two other common cases: (a) cluster-level assignments studies where the authors' model did not properly adjust for clustering (see equation E.7.0 in Table 2) and (b) individual-level assignment studies (see left-hand side for equation E.7.1 in Table 2).

This approach can serve as an alternative to standard error formulas based on $R^2$ values (equation E.2.2), which are often not reported. The following Section E addresses this alternate $R^2$ approach. If both the regression coefficient standard error and model $R^2$ value are reported, the WWC will use the coefficient standard error approach (Table 2). Finally, if the $t$-test statistic for an unstandardized coefficient is reported instead, the standard error can be calculated using $SE = b / t$.

If none of these uncertainty statistics are reported, the WWC will assume unadjusted effect sizes when computing the standard error (using equation E.1.4 for individual-level assignment studies and equation E.5.2 for cluster-level assignment studies), while still using the coefficient value to calculate the adjusted effect size.

**Table 2. Effect size standard error formulas based on regression model coefficients**

| Eq. number | Did the original study cluster correct? | Individual assignment | Cluster assignment |
|---|---|---|---|
| E.7.0 | Yes | Not applicable | $\omega\sqrt{\left(\frac{SE_{CC}}{S}\right)^2 \gamma + \frac{g^2}{2h}}$ |
| E.7.1 | No | $\omega\sqrt{\left(\frac{SE}{S}\right)^2 + \frac{g^2}{2N}}$ | $\omega\sqrt{\left(\frac{SE_{UC}}{S}\right)^2 \eta + \frac{g^2}{2h}}$ |

*Note.* $\eta = 1 + (n-1)\rho_{ICC}$; $\gamma = 1 - \frac{2(n-1)\rho_{ICC}}{N-2}$; $h = \frac{[(N-2)-2(n-1)\rho_{ICC}]^2}{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}^2+2(N-2n)\rho(1-\rho_{ICC})}$; $\rho_{ICC}$ is the intraclass correlation coefficient; $S$ is the pooled standard deviation; $SE_{CC}$ and $SE_{UC}$ are the regression coefficient standard errors corrected and uncorrected for clustering, respectively; $N = n_i + n_c$, or the total sample of individuals; $n$ is the average number of individuals per cluster.

# E. Other standard error formula extensions

For unadjusted effect sizes, equation E.5.2 of the *WWC Procedures Handbook* presents a general approach to compute the effect size standard error for cluster-assignment designs without covariates. This equation can be used when any of the following information is reported: unadjusted means and standard deviations, $t$-test for unadjusted means, and analysis of variance (ANOVA) $F$ statistics for unadjusted means.

For adjusted effect sizes, the following cluster-level assignment extensions in Table 3 provide the effect size standard errors for analysis of covariance (ANCOVA), OLS models, gain score difference-in-differences, and effect size difference-in-differences.

This table also introduces an extension to the standard error formulas for effect size differences-in-differences that is new for even individual-level assignment studies (see equation E.7.2). This new formula applies when (a) the effect size difference-in-differences adjustment is applied, and (b) an estimate of the baseline-outcome correlation $\rho_{cor}$ is provided. In contrast, the existing equation E.3.3 in Appendix E and its new cluster extension only apply when imputing default correlations (that is, a value of 1.0 in estimating the effect size and a value of 0.5 in estimating its variance), when an estimate of $\rho_{cor}$ is not provided.

Two versions for the effect-size difference-in-differences standard error formulas are needed to accommodate the two versions of relevant estimators. When the baseline-outcome correlation $\rho_{cor}$ is not reported, the difference-in-difference estimator is $g_{post} - g_{pre}$ based on equation E.3.2, and the standard error of this estimator is given by inserting 0.5 as the correlation in equation E.3.3. In contrast, when an estimate of $\rho_{cor}$ is provided, the difference-in-differences estimator becomes $g_{post} - \rho_{cor} g_{pre}$. Multiplying the correlation $\rho_{cor}$ to the baseline effect size $g_{pre}$ reduces the variance of the estimator, thereby requiring a different standard error formula (equation E.7.2).

**Table 3. Other extensions to standard error formulas**

| Eq. number | Calculation type | Individual assignment | Cluster assignment extensions |
|---|---|---|---|
| E.2.2 | ANCOVA adjusted means | $\omega\sqrt{\frac{n_i+n_c}{n_i n_c}(1-R^2)+\frac{g^2}{2(n_i+n_c)}}$ | $\omega\sqrt{\frac{n_i+n_c}{n_i n_c}(1-R^2)\eta+\frac{g^2}{2h}}$ |
| E.3.1 | Gain score DnD | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})+\frac{g^2}{2(n_i+n_c)}}$ | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})\eta+\frac{g^2}{2h}}$ |
| E.3.3 | Effect size DnD (no estimate of $\rho_{cor}$ given—impute as 0.5) | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})+\frac{{g_{post}}^2+{g_{pre}}^2-2g_{pre}g_{post}{\rho_{cor}}^2}{2(n_i+n_c)}}$ | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)2(1-\rho_{cor})\eta+\frac{{g_{post}}^2+{g_{pre}}^2-2g_{pre}g_{post}{\rho_{cor}}^2}{2h}}$ |
| E.7.2 | Effect size DnD (with estimate of $\rho_{cor}$ given) | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)(1-{\rho_{cor}}^2)+\frac{{g_{post}}^2+{g_{pre}}^2{\rho_{cor}}^2-2g_{pre}g_{post}{\rho_{cor}}^3}{2(n_i+n_c)}}$ | $\omega\sqrt{\left(\frac{n_i+n_c}{n_i n_c}\right)(1-{\rho_{cor}}^2)\eta+\frac{{g_{post}}^2+{g_{pre}}^2{\rho_{cor}}^2-2g_{pre}g_{post}{\rho_{cor}}^3}{2h}}$ |
| E.4.4 | Dichotomous outcomes | $\frac{1}{1.65}\sqrt{\frac{1}{p_i n_i}+\frac{1}{(1-p_i)n_i}+\frac{1}{p_c n_c}+\frac{1}{(1-p_c)n_c}}$ | $\frac{\sqrt{\eta}}{1.65}\sqrt{\frac{1}{p_i n_i}+\frac{1}{(1-p_i)n_i}+\frac{1}{p_c n_c}+\frac{1}{(1-p_c)n_c}}$ |

*Note.* $\eta = 1 + (n-1)\rho_{ICC}$; $h = \frac{[(N-2)-2(n-1)\rho_{ICC}]^2}{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}^2+2(N-2n)\rho_{ICC}(1-\rho_{ICC})}$; $\rho_{ICC}$ is the intraclass correlation coefficient; $\rho_{cor}$ is the baseline-outcome correlation; $N = n_i + n_c$, or the total sample of individuals; $n$ is the average number of individuals per cluster; $p_i$ and $p_c$ are the probabilities of an outcome occurring in the intervention and comparison groups, respectively. DnD = difference-in-differences.

# F. Errata

This supplement also corrects two minor errors in the original Version 4.1 standard error formulas (see red terms in Table 5), for unadjusted cluster-assignment effect size standard errors (equation E.5.2) and the effect size difference-in-differences standard errors (equation E.3.3).

**Table 4. Corrections to existing standard error formulas**

| Eq. number | Version | Formula |
|---|---|---|
| E.5.2 | Current formula | $\omega\sqrt{\left(\frac{N_i+N_c}{N_iN_c}\right)(1+(n-1)\rho_{ICC})+g^2\frac{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}{}^2+2(N-2n)\rho_{ICC}(1-\rho_{ICC})}{2(N-2)[(N-2)-2(n-1)\rho_{ICC}]}}$ |
| | Corrected formula | $\omega\sqrt{\left(\frac{N_i+N_c}{N_iN_c}\right)(1+(n-1)\rho_{ICC})+g^2\frac{(N-2)(1-\rho_{ICC})^2+n(N-2n)\rho_{ICC}{}^2+2(N-2n)\rho_{ICC}(1-\rho_{ICC})}{2[(N-2)-2(n-1)\rho_{ICC}]^2}}$ |
| E.3.3 | Current formula | $\omega\sqrt{\left(\frac{n_i+n_c}{n_in_c}\right)2(1-\rho_{cor})+\frac{g_{post}{}^2+g_{pre}{}^2+2g_{pre}g_{post}\rho_{cor}{}^2}{2(n_i+n_c)}}$ |
| | Corrected formula | $\omega\sqrt{\left(\frac{n_i+n_c}{n_in_c}\right)2(1-\rho_{cor})+\frac{g_{post}{}^2+g_{pre}{}^2-2g_{pre}g_{post}\rho_{cor}{}^2}{2(n_i+n_c)}}$ |

*Note.* $\eta = 1 + (n-1)\rho_{ICC}$; $\gamma = 1 - \frac{2(n-1)\rho_{ICC}}{N-2}$ ; $\rho_{cor}$ is the baseline-outcome correlation; $\rho_{ICC}$ is the intraclass correlation coefficient; $N = n_i + n_c$, or the total sample of individuals; $n$ is the average number of individuals per cluster.

## G. Implications for WWC Study Reviewers

Most of these extensions affect only back-end calculations for the Online Study Review Guide (OSRG). For example, this supplement introduces no changes in how study reviewers should enter and use study-reported *p*-values. However, the minor changes for OSRG users are:

1. If an unstandardized regression coefficient is used to compute the effect size, reviewers should enter the coefficient's standard error or *t*-test statistic if the study reported it.
2. If both the coefficient standard error and model $R^2$ value is reported, study reviewers should enter both values, although the OSRG will prioritize using the coefficient standard error.
3. For cluster-assignment designs, reviewers must indicate if the original study authors' regression model did or did not adjust for clustering.

This reviewer guidance will be incorporated into a forthcoming Version 4.1 OSRG User's Guide that will serve as the authoritative and most up to date OSRG reviewer guidance document.