Package: baselinr (via r-universe)

October 30, 2026

Title What Works Clearinghouse Standards for Education Impact Evaluations

Version 0.6.0.9000

Description Applies the group-design determinations of the What Works Clearinghouse (WWC) to education impact studies. Computes WWC effect sizes (Hedges' g with the small-sample correction, and the Cox index) and classifies baseline equivalence; classifies overall and differential attrition against the WWC attrition boundary; returns the group-design study rating; and reports the robustness of the baseline-equivalence verdict as a multiverse over the computation choices an analyst could defensibly make differently. Provides report-ready tables and Love plots.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Roxygen list(markdown = TRUE)

Imports rlang, stats

Suggests ggplot2, gt, knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

URL https://github.com/zl1212-ship-it/baselinr, https://zl1212-ship-it.github.io/baselinr/

BugReports https://github.com/zl1212-ship-it/baselinr/issues

Config/roxygen2/version 8.0.0

Depends R (>= 3.5)

LazyData true

Repository https://zl1212-ship-it.r-universe.dev

Date/Publication 2026-09-30 08:18:13 UTC

RemoteUrl https://github.com/zl1212-ship-it/baselinr

RemoteRef HEAD

RemoteSha 0bf173bc9e6142e81b404a1690f353581264bb0b

Contents

attrition . . . 2
attrition_boundary . . . 3
baseline_equivalence . . . 4
cluster_correction . . . 6
cox_index . . . 7
gt_baseline . . . 8
hedges_g . . . 9
love_plot . . . 10
tutoring . . . 11
wwc_classify . . . 11
wwc_rating . . . 12
wwc_robustness . . . 13
wwc_summary . . . 14

Index **16**

`attrition` *Overall and differential attrition*

Description

Computes overall and differential sample attrition for a two-group design, the inputs to the What Works Clearinghouse (WWC) attrition standard. Differential attrition is the absolute difference between the treatment and comparison attrition rates.

Usage

```
attrition(treatment, retained, na.rm = TRUE)
```

Arguments

`treatment`	Vector identifying group membership; exactly two unique non-missing values (the larger is treated as the treatment group, as in `hedges_g()`).
`retained`	Logical (or `0`/`1`) the same length as `treatment`: `TRUE`/`1` for cases retained in the analytic sample, `FALSE`/`0` for those lost.
`na.rm`	Logical; drop rows where `treatment` or `retained` is `NA`. Default `TRUE`.

Details

This function reports the attrition rates; it does not classify them. Compare the overall and differential rates against the WWC attrition boundary for your chosen response assumption (cautious or optimistic) in the Procedures Handbook.

Value

A one-row data frame with columns `attrition_overall`, `attrition_treatment`, `attrition_comparison`, and `differential_attrition` (all proportions).

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
set.seed(1)
g <- rep(c(1, 0), each = 100)
kept <- rbinom(200, 1, ifelse(g == 1, 0.9, 0.8))
attrition(g, kept)
```

`attrition_boundary` *Classify a study under the WWC attrition standard*

Description

Given a study's overall and differential attrition, classifies it as low or high attrition against the What Works Clearinghouse (WWC) attrition boundary (Standards Handbook Version 4.1, Table II.1). This is the classification that `attrition()` deliberately leaves to the user: `attrition()` reports the rates, `attrition_boundary()` applies the standard.

Usage

```
attrition_boundary(
  overall,
  differential,
  assumption = c("cautious", "optimistic")
)
```

Arguments

`overall`	Overall attrition, as a proportion in `[0, 1]` (as returned by `attrition()`).
`differential`	Differential attrition, as a proportion (the absolute difference in group attrition rates, as returned by `attrition()`).
`assumption`	Which boundary to apply: `"cautious"` (default) or `"optimistic"`.

Details

The WWC uses one of two boundaries. The **cautious** boundary is applied when the intervention could plausibly affect attrition (for example, a dropout prevention program); the **optimistic** boundary when it is unlikely to (for example, a first-grade reading program). The applicable boundary is set by the review protocol, not chosen post hoc; the default here is the more conservative cautious boundary.

Value

A data frame, one row per input, with columns `overall`, `differential`, `assumption`, `max_differential` (the highest differential attrition still counted as low, as a proportion; `NA` where the overall rate is beyond the boundary), and `attrition` (`"low"` or `"high"`).

References

What Works Clearinghouse (2020). *Standards Handbook, Version 4.1*, Table II.1. U.S. Department of Education.

Examples

```
# A study with 8% overall and 3 percentage-point differential attrition:
attrition_boundary(0.08, 0.03) # low under the cautious boundary

# 30% overall, 6-point differential: high if cautious, low if optimistic
attrition_boundary(0.30, 0.06, "cautious")
attrition_boundary(0.30, 0.06, "optimistic")

# Chained from attrition():
set.seed(1)
g <- rep(c(1, 0), each = 100)
kept <- rbinom(200, 1, ifelse(g == 1, 0.9, 0.82))
a <- attrition(g, kept)
attrition_boundary(a$attrition_overall, a$differential_attrition)
```

---

baseline_equivalence *Baseline equivalence table for an impact evaluation*

Description

Builds a report-ready baseline-equivalence table for a set of covariates, reporting group sample sizes, summaries, the appropriate standardized effect size, and the corresponding What Works Clearinghouse (WWC) equivalence category for each covariate. Continuous covariates use Hedges' g; binary covariates use the Cox index.

Usage

```
baseline_equivalence(
  data,
  treatment,
  covariates = NULL,
  missing = c("pairwise", "complete"),
  report_missing = FALSE
)
```

Arguments

data	A data frame.
treatment	String naming the column in data that identifies group membership. Must have exactly two unique non-missing values (see hedges_g() for how the treatment group is determined).
covariates	Character vector of column names to evaluate. Defaults to all numeric, logical, and factor columns in data other than treatment.
missing	How to handle missing covariate values. "pairwise" (the default, and the behavior of earlier versions) computes each covariate's row from the cases non-missing on that covariate and on treatment, so rows can describe different subsets. "complete" restricts every row to the cases complete on treatment and on all evaluated covariates, so every row describes the same set of cases.
report_missing	Logical; if TRUE, adds missing_treatment and missing_comparison columns reporting each group's proportion of missing values on the covariate. The rates are computed on all cases with a non-missing treatment value, before any deletion, under either setting of missing. Default FALSE.

Details

A covariate with exactly two unique non-missing values is treated as binary; any other numeric covariate is treated as continuous. A non-numeric covariate with more than two categories is not supported and raises an error. The binary-versus-continuous decision uses the sample actually contributing to the row, so it can differ between missing = "pairwise" and missing = "complete" when deletion removes all cases at one value.

Under WWC review, baseline equivalence must be established on the analytic sample, the cases actually contributing to the impact estimate. This function sees only the covariates, never the outcome, so neither option identifies the analytic sample by itself; to establish equivalence on the analytic sample, pass that subset of cases as data. Within the data supplied, missing = "complete" keeps every row on one common set of cases, as a complete-case analysis would, while "pairwise" uses all available information per covariate. report_missing = TRUE documents how much is missing in each group, complementing the sample-loss reporting in attrition().

Value

A data frame with one row per covariate and the columns: covariate; type ("continuous" or "binary"); n_treatment, n_comparison; mean_treatment, mean_comparison (group means for continuous covariates, event proportions for binary ones); sd_treatment, sd_comparison; effect_size (Hedges' g or Cox index, per type); and wwc_category. If report_missing = TRUE, the columns missing_treatment and missing_comparison appear after the sample-size columns.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
df <- data.frame(
  treat = c(1, 1, 1, 0, 0, 0),
  pretest = c(5, 6, 7, 4, 5, 6),
  female = c(1, 0, 1, 0, 0, 1)
)
baseline_equivalence(df, treatment = "treat")
```

`cluster_correction` *WWC clustering correction for mismatched analyses*

Description

Corrects the statistical significance of a finding for clustering, following the What Works Clearinghouse (WWC) procedure based on Hedges (2007). The correction applies when assignment was at the cluster level (classrooms or schools) but the analysis was at the student level, a "mismatch" that leaves the reported standard errors too small. It computes the t statistic implied by the effect size, then corrects both the t statistic and its degrees of freedom for clustering, and returns the clustering-corrected two-tailed p value and significance decision.

Usage

```
cluster_correction(
  g,
  n_treatment,
  n_comparison,
  m_treatment,
  m_comparison,
  icc = NULL,
  outcome = c("achievement", "other")
)
```

Arguments

`g`	Effect size (Hedges' g) ignoring clustering.
`n_treatment`, `n_comparison`	Student-level sample sizes.
`m_treatment`, `m_comparison`	Number of clusters in each group.
`icc`	Intraclass correlation. If `NULL` (default), the WWC default is used: 0.20 for `outcome = "achievement"`, 0.10 otherwise.
`outcome`	`"achievement"` or `"other"`, selecting the default ICC when `icc` is `NULL`.

Details

Because the correction can only reduce significance, the WWC applies it only to findings the study authors reported as statistically significant; a finding that was not significant stays not significant. This function computes the corrected values regardless; apply them where the WWC rules call for it.

Value

A data frame, one row per input, with columns `t` (ignoring clustering), `t_corrected`, `df` (corrected degrees of freedom), `p_value` (clustering-corrected, two-tailed), and `significant` (`p_value < 0.05`).

References

What Works Clearinghouse (2020). *Procedures Handbook, Version 4.1*, Appendix F. Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341-370.

Examples

```
# A finding with a moderate effect from a clustered design:
cluster_correction(
  g = 0.30, n_treatment = 200, n_comparison = 200,
  m_treatment = 10, m_comparison = 10, outcome = "achievement"
)
```

---

cox_index *Cox index for a binary covariate*

Description

Computes the What Works Clearinghouse (WWC) Cox index, a standardized effect size for a binary (dichotomous) covariate. The Cox index places the difference between two proportions on a scale comparable to Hedges' g, so it can be classified with the same baseline-equivalence thresholds.

Usage

```
cox_index(x, treatment, na.rm = TRUE)
```

Arguments

`x`	A binary covariate (numeric `0/1`, logical, two-level factor, or any vector with exactly two unique non-missing values). The larger value (e.g. `1`, `TRUE`, or the second sorted level) is treated as the "event".
`treatment`	Vector the same length as `x` identifying group membership; exactly two unique non-missing values (see `hedges_g()`).
`na.rm`	Logical; drop rows where `x` or `treatment` is `NA`. Default `TRUE`. When `FALSE`, missing values are not allowed: the function errors if any are present.

Details

The index is $d_{Cox} = (\text{logit}(p_t) - \text{logit}(p_c))/1.65$, where p_t and p_c are the proportions in the "event" category for the treatment and comparison groups.

Value

A single numeric value: the Cox index. Returns `NA` (with a warning) when a group proportion is exactly 0 or 1, where the index is undefined.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
x <- c(1, 1, 1, 1, 0, 1, 0, 0)
g <- c(1, 1, 1, 1, 0, 0, 0, 0)
cox_index(x, g)
```

---

gt_baseline *Format a baseline equivalence table with gt*

Description

Renders the result of `baseline_equivalence()` as a formatted **gt** table with rounded statistics and readable column labels. Requires the **gt** package.

Usage

```
gt_baseline(equivalence, decimals = 2)
```

Arguments

`equivalence`	A data frame returned by `baseline_equivalence()`.
`decimals`	Number of decimal places for the numeric columns. Default 2.

Value

A `gt_tbl` object.

Examples

```
if (requireNamespace("gt", quietly = TRUE)) {
  df <- data.frame(
    treat = c(1, 1, 1, 0, 0, 0),
    pretest = c(5, 6, 7, 4, 5, 6),
    female = c(1, 0, 1, 0, 0, 1)
  )
  tbl <- gt_baseline(baseline_equivalence(df, "treat"))
}
```

---

hedges_g *Hedges' g standardized mean difference*

Description

Computes the standardized mean difference (Hedges' g) between a treatment and a comparison group for a single numeric covariate, using the pooled within-group standard deviation and the small-sample correction factor used by the What Works Clearinghouse (WWC).

Usage

```
hedges_g(x, treatment, na.rm = TRUE)
```

Arguments

`x`	Numeric vector of covariate values.
`treatment`	Vector the same length as `x` identifying group membership. Must have exactly two unique non-missing values. The larger value (e.g. `1`, `TRUE`, or the second sorted level) is treated as the treatment group; the other as the comparison group.
`na.rm`	Logical; drop rows where `x` or `treatment` is `NA`. Default `TRUE`. When `FALSE`, missing values are not allowed: the function errors if any are present.

Details

The correction factor is $\omega = 1 - 3/(4N - 9)$, where $N = n_{treatment} + n_{comparison}$.

Value

A single numeric value: Hedges' g. Positive when the treatment group mean exceeds the comparison group mean.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
x <- c(5, 6, 7, 4, 5, 6)
g <- c(1, 1, 1, 0, 0, 0)
hedges_g(x, g) # 0.8
```

`love_plot` *Love plot of standardized effect sizes*

Description

Plots the standardized effect size for each covariate from `baseline_equivalence()`, with reference lines at the What Works Clearinghouse (WWC) thresholds (0.05 and 0.25) and points coloured by WWC category. Requires the **ggplot2** package.

Usage

```
love_plot(equivalence, signed = FALSE)
```

Arguments

`equivalence`	A data frame returned by `baseline_equivalence()`.
`signed`	Logical. If `FALSE` (default), plot absolute effect sizes with reference lines at 0.05 and 0.25. If `TRUE`, plot signed effect sizes with symmetric reference lines and a line at zero.

Value

A `ggplot` object.

Examples

```
if (requireNamespace("ggplot2", quietly = TRUE)) {
  df <- data.frame(
    treat = c(1, 1, 1, 0, 0, 0),
    pretest = c(5, 6, 7, 4, 5, 6),
    female = c(1, 0, 1, 0, 0, 1)
  )
  love_plot(baseline_equivalence(df, "treat"))
}
```

tutoring *Simulated tutoring program evaluation*

Description

A small, **simulated** (not real) dataset for demonstrating baseline equivalence assessment in a quasi-experimental education evaluation. It represents 400 students: 200 who received a tutoring program and 200 comparison students who did not, with baseline covariates measured before the program and an outcome measured after. The treatment group is mildly positively selected, so the covariates span all three What Works Clearinghouse equivalence categories.

Usage

```
tutoring
```

Format

A data frame with 400 rows and 8 variables:

treat Treatment indicator: 1 = received tutoring, 0 = comparison.

pretest Baseline reading score (continuous).

attendance Baseline attendance rate, 0-1 (continuous).

age Age in years at baseline (continuous).

female 1 = female, 0 = not (binary).

frpl Eligible for free or reduced-price lunch: 1 = yes (binary).

ell English language learner: 1 = yes (binary).

posttest Reading score after the program (continuous outcome).

Source

Simulated for package examples with `data-raw/tutoring.R`; not real student data.

wwc_classify *Classify baseline equivalence under WWC standards*

Description

Maps standardized effect sizes to the three What Works Clearinghouse baseline-equivalence categories. Sign is ignored; classification uses the absolute value of the effect size.

Usage

```
wwc_classify(es)
```

Arguments

es	Numeric vector of standardized effect sizes (e.g. values returned by hedges_g() or cox_index()).

Value

A character vector the same length as es:

- "satisfied" when |es| <= 0.05 (no adjustment needed),
- "satisfied_with_adjustment" when 0.05 < |es| <= 0.25 (equivalence holds only if the covariate is adjusted for in the impact model),
- "not_satisfied" when |es| > 0.25. NA inputs return NA.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
wwc_classify(c(0.03, 0.12, 0.80))
```

---

wwc_rating *WWC group-design study rating*

Description

Applies the What Works Clearinghouse (WWC) group-design rating logic to the two determinations this package supports, sample attrition and baseline equivalence, and returns the study's rating. This encodes the main decision path of the WWC Standards Handbook (Version 4.1, Section II): a randomized controlled trial (RCT) with low attrition can meet standards without reservations; an RCT with high attrition is held to the same baseline requirement as a quasi-experimental design (QED); and a QED, or a high attrition RCT, meets standards with reservations only if baseline equivalence is established, and otherwise does not meet standards.

Usage

```
wwc_rating(design, baseline, attrition = NULL)
```

Arguments

design	"rct" or "qed".
baseline	The overall baseline-equivalence verdict, one of "satisfied", "satisfied_with_adjustment", or "not_satisfied" (the overall value from wwc_summary()).
attrition	For an RCT, "low" or "high" (the attrition value from attrition_boundary()); required for design = "rct", ignored for a QED.

Details

This function covers the attrition-and-equivalence path only. A real WWC review also checks that random assignment was not compromised, that there are no confounding factors, and that the required baseline measures were used; those judgments are the reviewer's and are assumed satisfied here.

Value

A length-one character string: one of `"Meets WWC Group Design Standards Without Reservations"`, `"Meets WWC Group Design Standards With Reservations"`, or `"Does Not Meet WWC Group Design Standards"`. The reasoning is attached as `attr(x, "basis")`.

References

What Works Clearinghouse (2020). *Standards Handbook, Version 4.1*, Section II. U.S. Department of Education.

Examples

```
# Low-attrition RCT: meets without reservations regardless of baseline.
wwc_rating("rct", baseline = "not_satisfied", attrition = "low")

# High-attrition RCT hinges on baseline equivalence, like a QED.
wwc_rating("rct", baseline = "satisfied_with_adjustment", attrition = "high")

# QED that fails baseline equivalence does not meet standards.
wwc_rating("qed", baseline = "not_satisfied")
```

---

wwc_robustness *Robustness of the WWC baseline-equivalence verdict*

Description

Reports how stable a baseline-equivalence verdict is to the computation choices a careful analyst might defensibly make differently. For each continuous covariate it recomputes the standardized difference under the cross of two choices, standardizing by the pooled versus the comparison-group standard deviation and applying the What Works Clearinghouse (WWC) small-sample correction or not, and records whether the covariate's WWC category changes. Binary covariates use the Cox index, which does not depend on these choices. It also reports whether the overall verdict changes.

Usage

```
wwc_robustness(data, treatment, covariates = NULL)
```

Arguments

data	A data frame.
treatment	String naming the treatment-indicator column (see baseline_equivalence()).
covariates	Character vector of covariate columns. Defaults to all eligible columns other than treatment.

Details

This is a multiverse, or specification-curve, view of a single WWC determination: it shows whether the verdict depends on which defensible choice is made.

Value

A data frame with one row per covariate and the columns covariate, type, category_default (the category under baselinr's default), the set of categories the covariate takes across the defensible choices, flips (whether that set has more than one category), and abs_es_min / abs_es_max (the range of the absolute effect size across choices). The overall verdict under each choice is attached as attr(x, "overall"), and attr(x, "overall_stable") is TRUE when the overall verdict is invariant.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education. Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712.

Examples

```
df <- data.frame(
  treat = c(1, 1, 1, 0, 0, 0),
  pretest = c(5, 6, 7, 4, 5, 6),
  female = c(1, 0, 1, 0, 0, 1)
)
r <- wwc_robustness(df, treatment = "treat")
r
attr(r, "overall")
```

---

wwc_summary *Overall WWC baseline-equivalence verdict*

Description

Summarizes a baseline_equivalence() table into a one-row overall assessment: how many covariates fall in each What Works Clearinghouse (WWC) category, the largest absolute effect size, and an overall verdict.

Usage

```
wwc_summary(equivalence)
```

Arguments

equivalence	A data frame returned by baseline_equivalence().

Details

The overall verdict follows the logic of the categories: if any covariate is "not_satisfied", baseline equivalence cannot be established ("not_satisfied"); otherwise, if any covariate requires adjustment, the verdict is "satisfied_with_adjustment" (equivalence holds only if those covariates are adjusted for in the impact model); otherwise "satisfied".

Value

A one-row data frame with columns n_covariates, n_satisfied, n_satisfied_with_adjustment, n_not_satisfied, max_abs_effect, and overall.

References

What Works Clearinghouse (2022). *Procedures and Standards Handbook* (Version 5.0). U.S. Department of Education.

Examples

```
df <- data.frame(
  treat = c(1, 1, 1, 0, 0, 0),
  pretest = c(5, 6, 7, 4, 5, 6),
  female = c(1, 0, 1, 0, 0, 1)
)
wwc_summary(baseline_equivalence(df, "treat"))
```

Index

∗ **datasets**
tutoring, 11

attrition, 2
attrition(), *3*, *5*
attrition_boundary, 3
attrition_boundary(), *12*

baseline_equivalence, 4
baseline_equivalence(), *8*, *10*, *14*, *15*

cluster_correction, 6
cox_index, 7
cox_index(), *12*

gt_baseline, 8

hedges_g, 9
hedges_g(), *2*, *5*, *7*, *12*

love_plot, 10

tutoring, 11

wwc_classify, 11
wwc_rating, 12
wwc_robustness, 13
wwc_summary, 14
wwc_summary(), *12*