Package: benchexcal (via r-universe)

May 18, 2026

Type Package

Title Benchmark, Expand, and Calibrate (BenchExCal) Trial Emulation Tools

Version 0.1.0

Description Lightweight tools for evaluating real-world evidence (RWE) studies that emulate randomized clinical trials (RCTs). Provides (1) computation of the three pre-specified RCT-DUPLICATE / ENCORE agreement metrics -- statistical significance agreement (SA), estimate agreement (EA), and standardized difference agreement (SD) -- and (2) the Benchmark, Expand, and Calibrate (BenchExCal) calibration of a Stage 2 RWE study using the divergence observed in a Stage 1 RCT-RWE pair, plus tipping-point sensitivity analysis and forest plots. Methods follow Wang et al. JAMA 2023;329:1376 and Wang et al. CPT 2025;117:1820.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.0)

Imports stats

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

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

Repository https://xxue064.r-universe.dev

Date/Publication 2026-05-18 02:21:33 UTC

RemoteUrl https://github.com/xxue064/benchexcal

RemoteRef HEAD

RemoteSha 1a2da2679a67b462d3b029668971f7fd24490e3b

Contents

`agreement` *Three RCT-DUPLICATE / ENCORE agreement metrics*

Description

For an RCT-database study pair (or RCT vs. emulation), computes the three pre-specified binary agreement metrics used in RCT-DUPLICATE (Wang et al., JAMA 2023) and ENCORE (Weberpals et al., CPT 2026): statistical significance agreement (SA), estimate agreement (EA), and standardized difference / SMD agreement (SD).

Usage

```
agreement(
  rct_hr,
  rct_lower,
  rct_upper,
  rwe_hr,
  rwe_lower,
  rwe_upper,
  trial_type = c("superiority", "noninferiority"),
  ni_margin = NULL
)
```

Arguments

`rct_hr`	Numeric. Point estimate (HR) from the RCT.
`rct_lower`, `rct_upper`	Numeric. 95% CI bounds for the RCT HR.
`rwe_hr`	Numeric. Point estimate (HR) from the database study.
`rwe_lower`, `rwe_upper`	Numeric. 95% CI bounds for the RWE HR.
`trial_type`	One of `"superiority"` (default) or `"noninferiority"`. If `"noninferiority"`, `ni_margin` should also be supplied to enable partial significance agreement (SAP).
`ni_margin`	Numeric on the HR scale (e.g. 1.3). The non-inferiority margin defined by the trial protocol. Only used when `trial_type = "noninferiority"`.

Details

Metric definitions follow Wang et al. (JAMA 2023):

- **SA (statistical significance agreement)**: TRUE when both point estimates AND both 95% CIs sit on the same side of the null (HR = 1). For non-inferiority trials, *partial* SA (SAP) is flagged if the database upper bound lies below the NI margin, even when the RCT was not powered for superiority.
- **EA (estimate agreement)**: TRUE when the database point estimate falls inside the RCT 95% CI.
- **SD (standardized difference)**: TRUE when |z| < 1.96, where

$$z = \frac{\log(\mathrm{HR}_{RCT}) - \log(\mathrm{HR}_{RWE})}{\sqrt{\mathrm{var}(\log \mathrm{HR}_{RCT}) + \mathrm{var}(\log \mathrm{HR}_{RWE})}}$$

  and each variance is derived from its 95% CI via SE = (log(UL) - log(LL)) / (2 * 1.96).

Value

An object of class `bx_agreement` (a list). Use `print()` or `summary()` for formatted output.

References

Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA*. 2023;329(16):1376-1385.

Weberpals J, Schneeweiss S, et al. Emulating Comparative Oncology Trials With Real-World Evidence Studies (ENCORE). *Clin Pharmacol Ther*. 2026.

Examples

```
# LEADER trial (Wang et al., JAMA 2023, Table 1, study #1)
agreement(
  rct_hr = 0.87, rct_lower = 0.78, rct_upper = 0.97,
  rwe_hr = 0.82, rwe_lower = 0.76, rwe_upper = 0.87
)

# PRONOUNCE (noninferiority, NI margin 1.225)
agreement(
  rct_hr = 1.28, rct_lower = 0.59, rct_upper = 2.79,
  rwe_hr = 1.35, rwe_lower = 0.94, rwe_upper = 1.93,
  trial_type = "noninferiority", ni_margin = 1.225
)
```

`calibrate` *BenchExCal calibration of a Stage 2 RWE estimate*

Description

Implements the Benchmark, Expand, Calibrate (BenchExCal) framework (Wang et al., *Clin Pharmacol Ther* 2025): uses the divergence observed between a Stage 1 RCT and its database emulation to calibrate the point estimate and 95% CI of a Stage 2 database study designed to inform a supplemental indication (no Stage 2 RCT yet).

Usage

```
calibrate(
  stage1_rct_hr,
  stage1_rct_lower,
  stage1_rct_upper,
  stage1_rwe_hr,
  stage1_rwe_lower,
  stage1_rwe_upper,
  stage2_rwe_hr,
  stage2_rwe_lower,
  stage2_rwe_upper,
  check_benchmark = TRUE
)
```

Arguments

`stage1_rct_hr`, `stage1_rct_lower`, `stage1_rct_upper`
: Stage 1 RCT HR and 95% CI.

`stage1_rwe_hr`, `stage1_rwe_lower`, `stage1_rwe_upper`
: Stage 1 database study HR and 95% CI (emulation of the Stage 1 RCT).

`stage2_rwe_hr`, `stage2_rwe_lower`, `stage2_rwe_upper`
: Stage 2 database study HR and 95% CI (the supplemental indication).

`check_benchmark`
: Logical. If `TRUE` (default), the Stage 1 RCT-RWE pair is evaluated with `agreement` and a warning is issued if all three agreement metrics fail (per BenchExCal, calibration should not proceed in that case).

Details

Let $\hat{\theta}_1$, $\hat{\theta}_1^*$, and $\hat{\theta}_2^*$ denote the Stage 1 RCT, Stage 1 RWE, and Stage 2 RWE log hazard ratios, with variances V_1, V_1^*, V_2^*. The Stage 1 divergence is

$$\hat{\xi}_1 = \hat{\theta}_1^* - \hat{\theta}_1.$$

Following Wang et al. (CPT 2025), the rescaled Stage 2 divergence is

$$\hat{\xi}_2 = \sqrt{V_2^*/V_1^*}\,\hat{\xi}_1,$$

and the calibrated Stage 2 log HR is

$$\hat{\theta}_{2,\mathrm{cal}} = \hat{\theta}_2^* - \hat{\xi}_2,$$

with variance inflated by the Stage 1 divergence uncertainty,

$$V_{2,\mathrm{cal}} = V_2^* + (V_1^* + V_1)\,\frac{V_2^*}{V_1^*}.$$

Sign convention. The paper writes the mean as $\hat{\theta}_2^* + \hat{\xi}_2$. Because $\xi = \theta^* - \theta$ (RWE minus RCT) is defined as the systematic bias, bias correction is a *subtraction*; this implementation uses $\hat{\theta}_2^* - \hat{\xi}_2$. If you would rather match the paper's formula literally, negate the returned `xi_hat_2` before applying.

Value

An object of class `bx_calibration`.

References

Wang SV, Russo M, Glynn RJ, et al. A Benchmark, Expand, and Calibration (BenchExCal) Trial Emulation Approach for Using Real-World Evidence to Support Indication Expansions. *Clin Pharmacol Ther*. 2025;117(6):1820-1828.

Examples

```
# Stage 1: LEADER liraglutide (RCT-DUPLICATE)
# Stage 2: hypothetical expanded indication, RWE only
cal <- calibrate(
  stage1_rct_hr = 0.87, stage1_rct_lower = 0.78, stage1_rct_upper = 0.97,
  stage1_rwe_hr = 0.82, stage1_rwe_lower = 0.76, stage1_rwe_upper = 0.87,
  stage2_rwe_hr = 0.85, stage2_rwe_lower = 0.78, stage2_rwe_upper = 0.93
)
print(cal)
```

`calibration_inputs` *Print the inputs required for BenchExCal calibration*

Description

A quick reference that prints what you need to supply to `calibrate`. Call this if you forget which six (or nine) HR/CI values `calibrate()` expects.

Usage

```
calibration_inputs()
```

Value

Invisibly returns `NULL`. Called for the printed side effect.

Examples

```
calibration_inputs()
```

`plot.bx_calibration` *Forest plot for a BenchExCal calibration*

Description

Plots the Stage 1 RCT, Stage 1 RWE, Stage 2 RWE (uncalibrated), and Stage 2 RWE (calibrated) HRs on a single log-scale forest plot.

Usage

```
## S3 method for class 'bx_calibration'
plot(x, ...)
```

Arguments

`x`	A `bx_calibration` object.
`...`	Currently unused.

Value

A `ggplot` object.

`plot.bx_tipping` *Tipping-point plot*

Description

Visualises how the Stage 2 calibrated HR (with CI) shifts as a function of the assumed Stage 2 divergence ξ_2.

Usage

```
## S3 method for class 'bx_tipping'
plot(x, ...)
```

Arguments

`x`	A `bx_tipping` object.
`...`	Currently unused.

Value

A `ggplot` object.

rct_duplicate *Selected RCT-DUPLICATE trial emulation results*

Description

Effect estimates (pooled, adjusted) from selected trials in Wang et al. (JAMA 2023), useful for testing and illustrating `agreement` and `calibrate`.

Usage

```
data(rct_duplicate)
```

Format

A data frame with the following columns:

trial Trial acronym.

rct_hr, rct_lower, rct_upper RCT HR and 95% CI.

rwe_hr, rwe_lower, rwe_upper Database study (adjusted, pooled) HR and 95% CI.

trial_type `"superiority"` or `"noninferiority"`.

ni_margin NI margin on the HR scale, or `NA`.

Source

Wang SV et al., JAMA 2023;329:1376-1385, Table 1.

tipping_point *Tipping point sensitivity analysis for BenchExCal calibration*

Description

Sweeps over a range of plausible Stage 2 divergence values ξ_2 (by default from the 5th to the 95th percentile of the BenchExCal prior on ξ_2) and reports the calibrated Stage 2 HR at each point. The *tipping point* is the value of ξ_2 at which the calibrated 95% CI just crosses the decision threshold (HR = 1 by default).

Usage

```
tipping_point(x, n_grid = 200, quantile_range = c(0.05, 0.95), threshold = 1)
```

Arguments

x	A bx_calibration object returned by calibrate.
n_grid	Integer. Number of grid points to evaluate (default 200).
quantile_range	Length-2 numeric. Quantile range of the ξ_2 prior to scan (default c(0.05, 0.95)).
threshold	Numeric on the HR scale. Decision threshold for the tipping point (default 1, i.e., the null).

Details

At each grid value of ξ_2, the function computes

$$\hat{\theta}_2(\xi_2) = \log(\hat{\mathrm{HR}}_2^*) - \xi_2$$

and constructs a 95% CI using the uncalibrated Stage 2 standard error (the prior uncertainty in ξ_2 is already represented by the breadth of the scan itself, so additional variance inflation would double-count). The first grid point at which the relevant CI bound crosses the threshold is returned as the tipping point.

Value

An object of class bx_tipping.

Index