

Package ‘stepwedgpower’

July 14, 2026

Type Package

Title Stepped-Wedge Clinical Trial Analysis and Power Simulation

Version 0.1.3

Description Provides reusable functions for aggregated cluster-period data, mixed-effects analysis, and simulation-based power and type I error evaluation in stepped-wedge cluster randomized trials. The design and mixed-effects analysis follow Hussey and Hughes (2007) [<doi:10.1016/j.cct.2006.05.007>](https://doi.org/10.1016/j.cct.2006.05.007). Intraclass correlations for binary outcomes are converted to logistic-normal random-intercept standard deviations following Eldridge, Ukoumunne and Carlin (2009) [<doi:10.1111/j.1751-5823.2009.00092.x>](https://doi.org/10.1111/j.1751-5823.2009.00092.x). Monte Carlo uncertainty in estimated power is summarized using the exact binomial interval of Clopper and Pearson (1934) [<doi:10.1093/biomet/26.4.404>](https://doi.org/10.1093/biomet/26.4.404). The simulation engine supports sequence-specific baseline risks, cluster random effects, direct intraclass-correlation specification, Monte Carlo uncertainty intervals, and model-fitting diagnostics. Applied physician and specialty helpers are retained for backward compatibility and for an example health-services workflow.

License MIT + file LICENSE

Encoding UTF-8

Imports lme4

Suggests testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Lin (Amanda) Li [aut, cre],
Florin Vaida [aut] (degree: PhD)

Maintainer Lin (Amanda) Li <amandali14124277@gmail.com>

Repository CRAN

Date/Publication 2026-07-13 23:00:02 UTC

Contents

stepwedgepower-package	2
analyze_lpa_outcomes	2
cluster_sd_to_icc	3
estimate_power	4
estimate_specialty_rates	5
estimate_type1_error	6
fit_specialty_rate_model	8
icc_to_cluster_sd	9
prepare_physician_data	9
read_example_physician_data	10
run_stepwedge_analysis	10
simulate_stepwedge_trial	11
summarize_by_specialty	13
Index	14

stepwedgepower-package	<i>stepwedgepower: Tools for stepped-wedge clinical trial analysis and power simulation</i>
------------------------	---

Description

The **stepwedgepower** package refactors a one-off academic analysis script into reusable tools for physician-level data preparation, specialty-level rate modeling, and simulation-based planning for stepped-wedge studies.

analyze_lpa_outcomes	<i>Reproduce the core Lp(a) outcome analyses</i>
----------------------	--

Description

Fits the main outcome models from the original script for both overall Lp(a) testing and Lp(a) testing among patients with elevated LDL.

Usage

```
analyze_lpa_outcomes(  
  data,  
  provider_var = "prov_id",  
  specialty_var = "specialty",  
  outcomes = list(  
    overall = list(successes = "n_lpa_pat", trials = "n_total_pat"),  
    high_ldl = list(successes = "n_ldl_lpa_pat", trials = "n_ldl_pat")  
  )  
)
```

```

    ),
    links = c("logit", "identity"),
    nAGQ = 10
  )

```

Arguments

data	A physician-level analysis data frame.
provider_var	Provider identifier column.
specialty_var	Specialty column.
outcomes	Named list defining success and trial columns for each outcome.
links	Character vector of links to fit.
nAGQ	Number of quadrature points for <code>glmer()</code> .

Value

A nested list containing fitted models and specialty-rate tables.

cluster_sd_to_icc	<i>Convert a random-intercept standard deviation to a logistic-model ICC</i>
-------------------	--

Description

Convert a random-intercept standard deviation to a logistic-model ICC

Usage

```
cluster_sd_to_icc(cluster_sd)
```

Arguments

cluster_sd	Nonnegative random-intercept standard deviation.
------------	--

Value

The corresponding latent-scale intraclass correlation coefficient.

Examples

```
cluster_sd_to_icc(0.416)
```

estimate_power	<i>Estimate power by repeated stepped-wedge simulation</i>
----------------	--

Description

Monte Carlo uncertainty is reported as an exact (Clopper-Pearson) interval, which retains nominal coverage when the estimated power is at or near the boundary.

Usage

```
estimate_power(
  n_simulations = 100,
  alpha = 0.05,
  treatment_or = 2,
  n_clusters_per_sequence = c(10, 10, 10, 10),
  sequence_names = paste0("Sequence ", seq_along(n_clusters_per_sequence)),
  cluster_sd = 1.21,
  icc = NULL,
  baseline_probs = c(0.07, 0.04, 0.03, 0.02),
  n_per_cluster_period = 20,
  n_periods = length(n_clusters_per_sequence) + 1L,
  fit_link = c("logit", "identity"),
  seed = NULL,
  nAGQ = 1,
  effect_size_or = NULL,
  n_providers_per_specialty = NULL,
  specialty_names = NULL,
  tau_provider = NULL,
  base_probs = NULL,
  pts_per_step = NULL,
  n_steps = NULL
)
```

Arguments

n_simulations	Number of simulations.
alpha	Significance threshold.
treatment_or	Odds ratio for treatment under the data-generating model.
n_clusters_per_sequence	Integer vector giving the number of clusters in each sequence.
sequence_names	Labels for the stepped-wedge sequences.
cluster_sd	Standard deviation of the cluster random intercept. Supply either this argument or icc.
icc	Optional latent-scale intraclass correlation coefficient for a logistic random-intercept model.

baseline_probs	Baseline outcome probabilities for each sequence.
n_per_cluster_period	Number of observations per cluster-period.
n_periods	Number of study periods. The default is one baseline period plus one crossover period per sequence.
fit_link	Link used when fitting the analysis model.
seed	Optional random seed.
nAGQ	Number of quadrature points for the fitted mixed model.
effect_size_or	Deprecated alias for treatment_or.
n_providers_per_specialty	Deprecated alias for n_clusters_per_sequence.
specialty_names	Deprecated alias for sequence_names.
tau_provider	Deprecated alias for cluster_sd.
base_probs	Deprecated alias for baseline_probs.
pts_per_step	Deprecated alias for n_per_cluster_period.
n_steps	Deprecated alias for n_periods.

Value

An object of class `stepwedge_power` containing estimated power, Monte Carlo uncertainty, fit diagnostics, and simulation p-values.

Examples

```
estimate_power(
  n_simulations = 20, treatment_or = 2,
  n_clusters_per_sequence = c(10, 10, 10, 10),
  baseline_probs = rep(0.05, 4), icc = 0.05,
  n_per_cluster_period = 20, seed = 1
)
```

```
estimate_specialty_rates
```

Estimate specialty-specific probabilities from a fitted model

Description

Extracts specialty-level probabilities from a fitted specialty model.

Usage

```
estimate_specialty_rates(
  model,
  specialty_levels = NULL,
  specialty_var = "specialty",
  link = c("logit", "identity"),
  approximate_marginal = TRUE,
  logit_scale_factor = 0.346
)
```

Arguments

model A fitted model returned by `fit_specialty_rate_model`.

specialty_levels Optional vector of specialty levels.

specialty_var Name of the specialty column used in the model.

link Link function for the fitted model.

approximate_marginal Logical; whether to apply the random-intercept logit approximation.

logit_scale_factor Approximation constant used in the shrinkage factor.

Value

A data frame with specialty-level linear predictors and probabilities.

estimate_type1_error	<i>Estimate type I error by repeated stepped-wedge simulation</i>
----------------------	---

Description

Equivalent to `estimate_power` with the treatment odds ratio set to 1.

Usage

```
estimate_type1_error(
  n_simulations = 100,
  alpha = 0.05,
  n_clusters_per_sequence = c(10, 10, 10, 10),
  sequence_names = paste0("Sequence ", seq_along(n_clusters_per_sequence)),
  cluster_sd = 1.21,
  icc = NULL,
  baseline_probs = c(0.07, 0.04, 0.03, 0.02),
  n_per_cluster_period = 20,
  n_periods = length(n_clusters_per_sequence) + 1L,
  fit_link = c("logit", "identity"),
)
```

```

    seed = NULL,
    nAGQ = 1,
    n_providers_per_specialty = NULL,
    specialty_names = NULL,
    tau_provider = NULL,
    base_probs = NULL,
    pts_per_step = NULL,
    n_steps = NULL
  )

```

Arguments

n_simulations	Number of simulations.
alpha	Significance threshold.
n_clusters_per_sequence	Integer vector giving the number of clusters in each sequence.
sequence_names	Labels for the stepped-wedge sequences.
cluster_sd	Standard deviation of the cluster random intercept. Supply either this argument or <code>icc</code> .
icc	Optional latent-scale intraclass correlation coefficient for a logistic random-intercept model.
baseline_probs	Baseline outcome probabilities for each sequence.
n_per_cluster_period	Number of observations per cluster-period.
n_periods	Number of study periods. The default is one baseline period plus one crossover period per sequence.
fit_link	Link used when fitting the analysis model.
seed	Optional random seed.
nAGQ	Number of quadrature points for the fitted mixed model.
n_providers_per_specialty	Deprecated alias for <code>n_clusters_per_sequence</code> .
specialty_names	Deprecated alias for <code>sequence_names</code> .
tau_provider	Deprecated alias for <code>cluster_sd</code> .
base_probs	Deprecated alias for <code>baseline_probs</code> .
pts_per_step	Deprecated alias for <code>n_per_cluster_period</code> .
n_steps	Deprecated alias for <code>n_periods</code> .

Value

A `stepwedge_power` object with `type1_error` added.

Examples

```
estimate_type1_error(
  n_simulations = 20,
  n_clusters_per_sequence = c(10, 10, 10, 10),
  baseline_probs = rep(0.05, 4), icc = 0.05,
  n_per_cluster_period = 20, seed = 1
)
```

```
fit_specialty_rate_model
```

Fit a specialty-level testing-rate model

Description

Fits either a binomial GLM or a provider-random-intercept binomial GLMM for aggregated success/trial data.

Usage

```
fit_specialty_rate_model(
  data,
  successes,
  trials,
  specialty_var = "specialty",
  provider_var = NULL,
  link = c("logit", "identity"),
  random_intercept = !is.null(provider_var),
  nAGQ = 10
)
```

Arguments

<code>data</code>	A data frame containing counts and grouping variables.
<code>successes</code>	Name of the success-count column.
<code>trials</code>	Name of the trial-count column.
<code>specialty_var</code>	Name of the specialty column.
<code>provider_var</code>	Optional provider identifier column.
<code>link</code>	Link function. Supported values are "logit" and "identity".
<code>random_intercept</code>	Logical; whether to include a provider random intercept.
<code>nAGQ</code>	Number of adaptive Gauss-Hermite quadrature points for <code>lme4::glmer()</code> .

Value

A fitted `glm` or `merMod` object.

icc_to_cluster_sd	<i>Convert a logistic-model ICC to a random-intercept standard deviation</i>
-------------------	--

Description

Uses the latent-variable approximation for a logistic random-intercept model, where the level-1 variance is $\pi^2 / 3$.

Usage

```
icc_to_cluster_sd(icc)
```

Arguments

icc Intraclass correlation coefficient in the open interval (0, 1).

Value

The corresponding random-intercept standard deviation.

Examples

```
icc_to_cluster_sd(0.05)
```

prepare_physician_data	<i>Prepare physician-level stepped-wedge analysis data</i>
------------------------	--

Description

Filters the input data to the specialties of interest, applies panel-size thresholds, removes extreme outliers, and sorts the output.

Usage

```
prepare_physician_data(
  data,
  specialties = c("CARDIOLOGY", "FAMILY MEDICINE", "INTERNAL MEDICINE", "NEUROLOGY"),
  min_patients = 100,
  max_patients = 10000,
  specialty_var = "specialty",
  patient_var = "n_total_pat",
  provider_name_var = "PROV_NAME"
)
```

Arguments

<code>data</code>	A data frame.
<code>specialties</code>	Character vector of specialties to keep.
<code>min_patients</code>	Minimum total number of patients required.
<code>max_patients</code>	Maximum total number of patients allowed.
<code>specialty_var</code>	Name of the specialty column.
<code>patient_var</code>	Name of the total-patient count column.
<code>provider_name_var</code>	Name of the provider name column used for ordering.

Value

A filtered and sorted data frame.

```
read_example_physician_data
```

Read the bundled example physician data

Description

Reads a small synthetic physician-level example dataset bundled with the package.

Usage

```
read_example_physician_data()
```

Value

A data frame.

```
run_stepwedge_analysis
```

Fit the stepped-wedge analysis model to a simulated dataset

Description

Accepts either the generic column names (`events`, `n`, `intervention`, `period`, `sequence_index`, `cluster_id`) or the legacy 0.1.0 column names (`n_positive`, `n_patients`, `treat`, `step`, `specialty_idx`, `PID`).

Usage

```
run_stepwedge_analysis(sim_data, fit_link = c("logit", "identity"), nAGQ = 1)
```

Arguments

sim_data	A data frame generated by <code>simulate_stepwedge_trial</code> or a compatible aggregated cluster-period dataset.
fit_link	Link function used in the fitted model.
nAGQ	Number of quadrature points for <code>lme4::glmer()</code> .

Value

A list with fitted model, treatment coefficient table, p-value, and convergence diagnostics.

Examples

```
sim <- simulate_stepwedge_trial(n_clusters_per_sequence = c(5, 5, 5, 5),
                               baseline_probs = rep(0.1, 4),
                               icc = 0.05, seed = 1)
run_stepwedge_analysis(sim)$p_value
```

`simulate_stepwedge_trial`

Simulate one stepped-wedge trial dataset

Description

Generates aggregated cluster-by-period binomial data for a stepped-wedge design in which cluster sequences cross over sequentially.

Usage

```
simulate_stepwedge_trial(
  treatment_or = 1.5,
  n_clusters_per_sequence = c(40, 40, 40, 40),
  sequence_names = paste0("Sequence ", seq_along(n_clusters_per_sequence)),
  cluster_sd = 1.21,
  icc = NULL,
  baseline_probs = c(0.06, 0.04, 0.03, 0.02),
  n_per_cluster_period = 20,
  n_periods = length(n_clusters_per_sequence) + 1L,
  seed = NULL,
  effect_size_or = NULL,
  n_providers_per_specialty = NULL,
  specialty_names = NULL,
  tau_provider = NULL,
  base_probs = NULL,
  pts_per_step = NULL,
  n_steps = NULL
)
```

Arguments

treatment_or	Odds ratio for treatment under the data-generating model.
n_clusters_per_sequence	Integer vector giving the number of clusters in each sequence.
sequence_names	Labels for the stepped-wedge sequences.
cluster_sd	Standard deviation of the cluster random intercept. Supply either this argument or <code>icc</code> .
icc	Optional latent-scale intraclass correlation coefficient for a logistic random-intercept model.
baseline_probs	Baseline outcome probabilities for each sequence.
n_per_cluster_period	Number of observations per cluster-period.
n_periods	Number of study periods. The default is one baseline period plus one crossover period per sequence.
seed	Optional random seed.
effect_size_or	Deprecated alias for <code>treatment_or</code> .
n_providers_per_specialty	Deprecated alias for <code>n_clusters_per_sequence</code> .
specialty_names	Deprecated alias for <code>sequence_names</code> .
tau_provider	Deprecated alias for <code>cluster_sd</code> .
base_probs	Deprecated alias for <code>baseline_probs</code> .
pts_per_step	Deprecated alias for <code>n_per_cluster_period</code> .
n_steps	Deprecated alias for <code>n_periods</code> .

Value

A data frame with one row per cluster-period combination. Generic column names are supplied together with legacy aliases for compatibility.

Examples

```
sim <- simulate_stepwedge_trial(
  treatment_or = 1.5,
  n_clusters_per_sequence = c(10, 10, 10, 10),
  baseline_probs = rep(0.05, 4),
  icc = 0.05,
  n_per_cluster_period = 20,
  seed = 1
)
head(sim[, c("cluster_id", "sequence", "period", "intervention", "n", "events")])
```

`summarize_by_specialty`*Summarize physician counts by specialty*

Description

Computes common summary statistics for one or more numeric variables within each specialty.

Usage

```
summarize_by_specialty(  
  data,  
  specialty_var = "specialty",  
  vars = c("n_total_pat", "n_ldl_pat"),  
  na.rm = TRUE  
)
```

Arguments

<code>data</code>	A data frame.
<code>specialty_var</code>	Name of the specialty column.
<code>vars</code>	Character vector of numeric variable names to summarize.
<code>na.rm</code>	Logical; whether to remove missing values.

Value

A data frame with one row per specialty-variable combination.

Index

`analyze_lpa_outcomes`, [2](#)
`cluster_sd_to_icc`, [3](#)
`estimate_power`, [4](#), [6](#)
`estimate_specialty_rates`, [5](#)
`estimate_type1_error`, [6](#)
`fit_specialty_rate_model`, [6](#), [8](#)
`icc_to_cluster_sd`, [9](#)
`prepare_physician_data`, [9](#)
`read_example_physician_data`, [10](#)
`run_stepwedge_analysis`, [10](#)
`simulate_stepwedge_trial`, [11](#), [11](#)
`stepwedgpower-package`, [2](#)
`summarize_by_specialty`, [13](#)