Package: bbssr (via r-universe)

October 28, 2026

Type Package

Title Blinded Sample Size Re-Estimation for Binary Endpoints

Version 3.0.0

Description Tools for blinded sample size re-estimation (BSSR) in two-arm clinical trials with binary endpoints, together with the exact power and sample size calculations that the re-estimation relies on, for two independent groups, for matched pairs and for stratified designs. For two independent groups seven statistical tests are implemented: Pearson chi-squared, Fisher exact, Fisher mid-p, Z-pooled exact unconditional, the exact unconditional test of Boschloo (1970) <doi:10.1111/j.1467-9574.1970.tb00104.x>, and the tests of non-inferiority and equivalence of Blackwelder (1982) <doi:10.1016/0197-2456(82)90024-1> and of Farrington and Manning (1990) <doi:10.1002/sim.4780091208>, with a margin on the scale of the risk difference or the risk ratio. Each test is available with a one-sided or a two-sided alternative, and the exact unconditional tests can be combined with the Berger-Boos procedure. For matched pairs the McNemar tests and the score test of Nam (1997) <doi:10.2307/2533508> are implemented, and for stratified designs the Cochran-Mantel-Haenszel test, Cochran's test with the sample size formulas of Woolson et al. (1986) <doi:10.2307/2530706> and Nam (1992) <doi:10.2307/2532298>, and the stratified exact conditional test of Jung (2014) <doi:10.1002/bimj.201300048>. The power, the distribution of the final sample size and the type I error rate of a re-estimation design are computed exactly, by simulation for three or more strata and as averages over simulated allocation sequences for random stratum and group sizes, the significance level can be adjusted so that the type I error rate is controlled, and the sample size can be re-estimated from the blinded data of a trial that is under way. Statistical methods based on Mehrotra et al. (2003) <doi:10.1111/1541-0420.00051>, Berger and Boos (1994) <doi:10.1080/01621459.1994.10476836>, Friede and Kieser (2004)

<doi:10.1002/pst.140>, Friede et al. (2007)
<doi:10.1002/bimj.200610373> and Kieser (2020)
<doi:10.1007/978-3-030-49528-2_21>.

License MIT + file LICENSE

URL https://gosukehommaex.github.io/bbssr/,
https://github.com/gosukehommaEX/bbssr

BugReports https://github.com/gosukehommaEX/bbssr/issues

Depends R (>= 3.5.0)

Imports fpCompare, ggplot2 (>= 3.4.0), Rcpp, stats, utils

LinkingTo Rcpp

Suggests testthat (>= 3.1.7), knitr, rmarkdown, Exact, exact2x2, microbenchmark

VignetteBuilder knitr

Encoding UTF-8

Language en-GB

Roxygen list(markdown = TRUE)

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

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

Date/Publication 2026-10-28 00:26:39 UTC

RemoteUrl https://github.com/gosukehommaex/bbssr

RemoteRef HEAD

RemoteSha 976931545a22d5628200e12f7584c3f1d7d8c875

Contents

BinaryAlphaAdjBSSR 4
BinaryBSSR 9
BinaryCondRejectBSSR 13
BinaryGridBSSR 17
BinaryPower 19
BinaryPowerBSSR 21
BinaryRR 26
BinarySampleSize 31
BinaryTypeIErrorBSSR 35
PairedAlphaAdjBSSR 40
PairedBSSR 44
PairedPower 46
PairedPowerBSSR 48
PairedRR 51
PairedSampleSize 54

PairedTypeIErrorBSSR 57
plot.bbssr_crp 60
plot.bbssr_grid 61
plot.bbssr_power 62
plot.bbssr_powerbssr 64
plot.bbssr_prr 65
plot.bbssr_ptie 67
plot.bbssr_rr 68
plot.bbssr_samplesize 69
plot.bbssr_sgrid 71
plot.bbssr_spowerbssr 72
plot.bbssr_stie 73
plot.bbssr_tie 74
print.bbssr_alphaadj 75
print.bbssr_bssr 76
print.bbssr_crp 77
print.bbssr_pbssr 77
print.bbssr_power 78
print.bbssr_powerbssr 79
print.bbssr_ppower 79
print.bbssr_ppowerbssr 80
print.bbssr_prr 81
print.bbssr_psamplesize 82
print.bbssr_ptie 82
print.bbssr_rr 83
print.bbssr_samplesize 84
print.bbssr_sbssr 84
print.bbssr_scrp 85
print.bbssr_spower 86
print.bbssr_spowerbssr 87
print.bbssr_srr 88
print.bbssr_ssamplesize 88
print.bbssr_stie 89
print.bbssr_tie 90
StratAlphaAdjBSSR 91
StratBSSR 95
StratCondRejectBSSR 98
StratGridBSSR 101
StratPower 104
StratPowerBSSR 106
StratRR 111
StratSampleSize 113
StratTypeIErrorBSSR 116
summary.bbssr_grid 120
summary.bbssr_powerbssr 121
summary.bbssr_spowerbssr 122

Index **124**

BinaryAlphaAdjBSSR *Adjusted Significance Level of a Blinded Sample Size Re-estimation Design*

Description

Lowers the nominal significance level of a two-arm trial with a binary endpoint and blinded sample size re-estimation (BSSR) until the largest type I error rate over the common response probability, or over the pooled response probability on the boundary of a non-inferiority hypothesis, does not exceed the target level alpha, and returns the adjusted level together with that of the fixed-sample design.

Usage

```
BinaryAlphaAdjBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard", "null.variance", "alternative.variance"),
  ss.Test = Test,
  ss.alpha = alpha,
  rounding = c("group", "friede-kieser", "total", "nearest"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  theta = seq(0, 1, by = 0.005),
  maximize = c("certified", "refined", "grid"),
  adjust = c("test", "both"),
  tol = 1e-08,
  step = 1e-05,
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, used to split the blinded pooled proportion into group-specific proportions
N1	Initial sample size of group 1
N2	Initial sample size of group 2
omega	Fraction of the initial sample size observed at the interim analysis. The interim size of group 2 is ceiling(omega N2) and that of group 1 is ceiling(r ceiling(omega N2)), so the interim analysis keeps the allocation ratio. Either omega or n.interim must be supplied
r	Allocation ratio to group 1
alpha	Target level of significance, which the type I error rate must not exceed
tar.power	Target power
Test	Type of statistical test of the final analysis. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
restricted	Logical. If TRUE, the re-estimated sample size is not allowed to fall below the initial sample size. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), 'less', 'two.sided', or 'equivalence' for the alternative that the effect lies between the margins of a test of equivalence, see margin
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference p1 - p2, 'RR' for the risk ratio p1 / p2 or 'OR' for the odds ratio
ss.method	How the sample size is re-estimated from the recovered proportions. Options: 'exact' (default), 'standard', 'null.variance' or 'alternative.variance', as in BinarySampleSize
ss.Test	Test whose exact power is used for the re-estimation when ss.method = 'exact'. Default is Test
ss.alpha	Level of significance used for the re-estimation when adjust = 'test'. Default is alpha. It is ignored when adjust = 'both'
rounding	How the re-estimated sample size is turned into whole numbers. Options: 'group' (default), 'friede-kieser', 'total' or 'nearest'. Only 'group' is available with ss.method = 'exact'. See Details
search	How the exact re-estimation chooses the sample size when ss.method = 'exact': 'crossing' (default), 'smallest' or 'stable', as in BinarySampleSize, without going beyond the largest size that N.max allows. Ignored by the other methods
search.limit	Limit of the size of group 2 examined by search = 'stable', as in BinarySampleSize. Default is c(2, 50)

N.min — Lower bound on the final total sample size, or NULL (default) for none beyond the interim total. It can be used to keep the patients who are already enrolled but not yet evaluated at the interim analysis

N.max — Upper bound on the final total sample size, or NULL (default) for none

n.interim — Interim sample sizes of group 1 and group 2, as a vector of length two. An alternative to omega

theta — Grid of common response probabilities at which the type I error rate is evaluated. Default is seq(0, 1, by = 0.005). With a non-zero margin, or a margin on the scale of the risk ratio, theta is the pooled response probability (r p1 + p2) / (1 + r) on the boundary of the null hypothesis, see Details

maximize — How the largest type I error rate at a level is located, as in BinaryTypeIErrorBSSR. With 'certified' (default) the adjusted level controls the type I error rate over the whole interval from the smallest to the largest value of theta, restricted with a margin to the values at which both response probabilities on the null boundary lie in the unit interval, see Details. With 'refined' and 'grid' the control is assessed on the grid

adjust — Which part of the design uses the adjusted level. 'test' (default) applies it to the final analysis only, so the re-estimation keeps using ss.alpha. 'both' applies it to the re-estimation as well

tol — Tolerance of the bisection, relative to alpha, a single value in (0, 1). The bisection is used when adjust = 'test' and, whatever adjust, for the fixed-sample design. Default is 1e-8

step — Decrement of the level in the search used when adjust = 'both'. Default is 1e-5

margin — Non-inferiority margin, on the scale given by margin.scale. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p1 - p2 <= -margin against p1 - p2 > -margin when alternative is 'greater', and p1 - p2 >= margin against p1 - p2 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio R0, and the null hypothesis is p1 / p2 <= R0 against p1 / p2 > R0 when alternative is 'greater', and p1 / p2 >= R0 against p1 / p2 < R0 when it is 'less'. A margin other than 0, and any margin on the scale of the risk ratio, requires Test = 'Blackwelder' or 'Farrington-Manning' and a one-sided alternative or 'equivalence', see BinaryRR. With a margin other than 0 on the scale of the risk difference the assumed and the true effects are risk differences (effect = 'RD'), and with a margin on the scale of the risk ratio they are risk ratios (effect = 'RR'). For alternative = 'equivalence' the margin is the pair c(lower, upper) of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis p1 - p2 <= -lower or p1 - p2 >= upper is tested against -lower < p1 - p2 < upper. On the scale of the risk ratio the pair satisfies 0 < lower < 1 < upper, a single ratio R gives the pair of R and 1 / R, the smaller as lower, and the null hypothesis is p1 / p2 <= lower or p1 / p2 >= upper. The null hypothesis is rejected when both one-sided tests reject, each at the level alpha

ref.pvalue	Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is FALSE. It applies to the final analysis, to the fixed-sample comparator and to the exact re-estimation
margin.scale	Scale of margin. Options: 'RD' (default) for the risk difference p1 - p2 or 'RR' for the risk ratio p1 / p2

Details

Following Kieser and Friede (2000) and Friede and Kieser (2004), the nominal level of the final analysis is lowered until the largest type I error rate over the common response probability, evaluated as in BinaryTypeIErrorBSSR, does not exceed alpha.

With adjust = 'test' the final sample sizes do not depend on the nominal level, and the rejection regions shrink as the level decreases, so the largest type I error rate is a non-decreasing step function of the level. The bisection locates, to within tol * alpha, the level above which the largest rate exceeds alpha. The package rejects a p-value at a level when it is below the level by more than a tolerance of about 1.49e-8, so every level between the largest p-value rejected at the level found plus this tolerance and the smallest p-value not rejected gives the same rejection regions, also when a p-value is rejected if it is below the level or at most equal to it. The reported alpha.adj is the largest value with at most six significant digits in this range, or with more digits if there is none, so the printed value can be used as it is. The same applies to the fixed-sample design.

With adjust = 'both' the re-estimated sample sizes change with the level as well, so the type I error rate need not be monotone in the level. The level is then lowered from alpha in steps of step until the type I error rate is controlled. The result is the first level among alpha - step, alpha - 2 * step, and so on, at which the type I error rate is controlled, and it can change with step. The levels that control the type I error rate need not form an interval, so a level below the result can fail, and a level above it that is not on this grid can pass. This search re-estimates the sample size at every step and can take much longer than the bisection. The fixed-sample design always uses the bisection, since its sample size does not depend on the level. The script reproduce-published.R in the folder system.file('reproduce', package = 'bbssr') reproduces the published values of Section 5 of Friede and Kieser (2004) and of Example 21.1 of Kieser (2020) only when the re-estimation also uses the adjusted level, as with adjust = 'both'.

The type I error rate can also be controlled within a confidence interval for the common response probability, or for the pooled response probability on the null boundary, instead of over its whole range (Kieser, 2020, Section 22.1.3). A value gamma between 0 and alpha is fixed in advance, and the final analysis uses a level at which the largest type I error rate over the 1 - gamma confidence interval computed from the data of the completed trial does not exceed alpha - gamma. The type I error rate is then controlled at alpha. This level is found when theta spans the interval and alpha is set to alpha - gamma. The interval is known only at the end of the trial, so the re-estimation keeps the nominal level, with adjust = 'test' and ss.alpha set to the nominal level. Neither search examines levels above alpha - gamma. The script reproduce-published.R reproduces the adjusted level of Example 23.1 of Kieser (2020), which uses the Clopper-Pearson interval with gamma = 1e-4.

With maximize = 'certified' the levels are assessed on the grid with the refinement of maximize = 'refined', and the level found is then certified as in BinaryTypeIErrorBSSR: it is accepted only if the upper bound of the type I error rate over the interval from the smallest to the largest value of theta does not exceed alpha. If the bound exceeds alpha, the grid has missed a value

of theta at which the level fails, or the largest rate lies within 1e-12 below alpha, and the search continues below that level. The bisection then certifies every level that passes the assessment on the grid, and the search of adjust = 'both' certifies every such level from the start. The adjusted level therefore controls the type I error rate over the whole interval. If a certification stops at its limit of 1e5 subdivisions, with a warning, the bound can exceed the largest rate by more than 1e-12, and the level found can be lower than necessary.

In both searches the type I error rate at a new level is first evaluated at the single value of theta where the largest rate of the last level that failed was found. If it exceeds alpha there, the level fails without the evaluation over the whole grid. The largest rate over theta is never below the rate at any one value, so such a level also fails on the largest rate. With maximize = 'grid' this value is a grid point, and every decision is that of the evaluation on the grid. With 'refined' and 'certified' it can lie between the grid points, where the grid can miss a rate above alpha, so a level can fail that the evaluation on the grid would accept. This can only lower the level found, and with 'certified' the level found is certified in either case.

Value

An object of class bbssr_alphaadj, a data frame with one row for the BSSR design and one for the fixed-sample design containing:

Design 'BSSR' or 'TRAD'

alpha Target level

max.TIE Largest type I error rate at the nominal level alpha

alpha.adj Adjusted nominal level, see Details. It is alpha when the nominal level already controls the type I error rate, and 0, with a warning, when no level examined by the bisection does. The search of adjust = 'both' for the BSSR design stops with an error instead when it reaches 0

max.TIE.adj Largest type I error rate at the adjusted level, 0 when alpha.adj is 0

theta.adj Common response probability, or pooled response probability on the null boundary, at which max.TIE.adj occurs, NA when alpha.adj is 0

max.TIE.bound Upper bound of the type I error rate at the nominal level with maximize = 'certified', otherwise NA

max.TIE.adj.bound Upper bound of the type I error rate at the adjusted level with maximize = 'certified', otherwise NA

For alternative = 'equivalence' the largest rates are taken over both boundaries of the null hypothesis, see BinaryTypeIErrorBSSR, and the column boundary.adj gives the boundary, 'lower' or 'upper', on which max.TIE.adj occurs.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Kieser M, Friede T (2000). Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine*, 19(7), 901-911.

Friede T, Kieser M (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics*, 3(4), 269-279.

Kieser M (2020). *Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials*. Springer, Cham.

See Also

BinaryTypeIErrorBSSR

Examples

```
BinaryAlphaAdjBSSR(
  Delta.A = 0.3, N1 = 39, N2 = 39, n.interim = c(20, 20), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = seq(0.01, 0.99, by = 0.01)
)
```

BinaryBSSR *Sample Size Re-estimation from Observed Blinded Interim Data*

Description

Re-estimates the sample size of an ongoing two-arm trial with a binary endpoint from the blinded data available at an interim analysis, and reports how many patients still have to be enrolled in each group during the second stage. Only the total number of patients and the total number of responders are required, so the treatment allocation remains concealed.

Usage

```
BinaryBSSR(
  n1,
  n2,
  S,
  Delta.A,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  N1 = NULL,
  N2 = NULL,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  effect = c("RD", "RR", "OR"),
```

```
  ss.method = c("exact", "standard", "null.variance", "alternative.variance"),
  ss.Test = Test,
  ss.alpha = alpha,
  rounding = c("group", "friede-kieser", "total", "nearest"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

n1	Number of patients of group 1 observed at the interim analysis
n2	Number of patients of group 2 observed at the interim analysis
S	Total number of responders observed at the interim analysis, pooled over both groups
Delta.A	Assumed treatment effect, on the scale given by `effect`, used to split the blinded pooled proportion into group-specific proportions
r	Allocation ratio to group 1 (i.e., allocation ratio of group 1:group 2 = r:1, r > 0)
alpha	Level of significance of the final analysis, for the alternative specified by `alternative`
tar.power	Target power
Test	Type of statistical test of the final analysis. Options: `'Chisq'`, `'Fisher'`, `'Fisher-midP'`, `'Z-pool'`, `'Boschloo'`, `'Blackwelder'` or `'Farrington-Manning'`
restricted	Logical. If `TRUE`, the re-estimated sample size is not allowed to fall below the planned sample size given by `N1` and `N2`. Default is `FALSE`
N1	Planned sample size of group 1. Required when `restricted` is `TRUE`, and when the recovered proportions coincide
N2	Planned sample size of group 2. Required when `restricted` is `TRUE`, and when the recovered proportions coincide
alternative	Direction of the alternative hypothesis. Options: `'greater'` (default), `'less'`, `'two.sided'`, or `'equivalence'` for the alternative that the effect lies between the margins of a test of equivalence, see `margin`
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: `'minlike'` (default), `'central'` or `'blaker'`
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
effect	Scale of `Delta.A`. Options: `'RD'` (default), `'RR'` or `'OR'`, as in BinaryPowerBSSR
ss.method	How the sample size is re-estimated. Options: `'exact'` (default), `'standard'`, `'null.variance'` or `'alternative.variance'`, as in BinarySampleSize

ss.Test — Test whose exact power is used for the re-estimation when ss.method = 'exact'. Default is Test

ss.alpha — Level of significance used for the re-estimation. Default is alpha

rounding — How the re-estimated sample size is turned into whole numbers. Options: 'group' (default), 'friede-kieser', 'total' or 'nearest', as in BinaryPowerBSSR

search — How the exact re-estimation chooses the sample size when ss.method = 'exact': 'crossing' (default), 'smallest' or 'stable', as in BinarySampleSize, without going beyond the largest size that N.max allows. Ignored by the other methods

search.limit — Limit of the size of group 2 examined by search = 'stable', as in BinarySampleSize. Default is c(2, 50)

N.min — Lower bound on the final total sample size, or NULL (default). It can be used to keep the patients who are already enrolled but not yet evaluated

N.max — Upper bound on the final total sample size, or NULL (default)

margin — Non-inferiority margin, on the scale given by margin.scale. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p1 - p2 <= -margin against p1 - p2 > -margin when alternative is 'greater', and p1 - p2 >= margin against p1 - p2 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio R0, and the null hypothesis is p1 / p2 <= R0 against p1 / p2 > R0 when alternative is 'greater', and p1 / p2 >= R0 against p1 / p2 < R0 when it is 'less'. A margin other than 0, and any margin on the scale of the risk ratio, requires Test = 'Blackwelder' or 'Farrington-Manning' and a one-sided alternative or 'equivalence', see BinaryRR. With a margin other than 0 on the scale of the risk difference the assumed and the true effects are risk differences (effect = 'RD'), and with a margin on the scale of the risk ratio they are risk ratios (effect = 'RR'). For alternative = 'equivalence' the margin is the pair c(lower, upper) of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis p1 - p2 <= -lower or p1 - p2 >= upper is tested against -lower < p1 - p2 < upper. On the scale of the risk ratio the pair satisfies 0 < lower < 1 < upper, a single ratio R gives the pair of R and 1 / R, the smaller as lower, and the null hypothesis is p1 / p2 <= lower or p1 / p2 >= upper. The null hypothesis is rejected when both one-sided tests reject, each at the level alpha

ref.pvalue — Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is FALSE

margin.scale — Scale of margin. Options: 'RD' (default) for the risk difference p1 - p2 or 'RR' for the risk ratio p1 / p2

Details

The blinded estimate of the pooled response probability is hat.p = S / (n1 + n2). For the risk difference, group-specific proportions are recovered as hat.p1 = hat.p + Delta.A / (1 + r) and hat.p2 = hat.p - r Delta.A / (1 + r), truncated to the unit interval. For the risk ratio and the

odds ratio they are the proportions with the pooled value `hat.p` and the ratio `Delta.A`. The sample size is then re-estimated as in `BinarySampleSize`.

Under the unrestricted rule and `rounding = 'group'`, the final size of group 2 is the larger of the re-estimated size and what has already been observed. Under the restricted rule it is raised to the planned size first, so the trial can only grow, and `N.min` and `N.max` add a further lower and upper bound on the total. The final size of group 1 is then `ceiling(r N2.final)`, so an imbalance already present at the interim is corrected by the remaining enrolment instead of being carried forward. Neither second-stage size is ever negative. The other rounding rules are described in `BinaryPowerBSSR`.

While `BinaryPowerBSSR` evaluates the operating characteristics of a BSSR design at the planning stage, this function is applied once, to the data of a trial that is under way.

Value

An object of class `bbssr_bssr`, a data frame with one row containing:

n1 Interim sample size of group 1

n2 Interim sample size of group 2

n Total interim sample size

S Total number of interim responders

hat.p Blinded estimate of the pooled response probability

hat.p1 Recovered response probability of group 1

hat.p2 Recovered response probability of group 2

N1.re Re-estimated total sample size of group 1

N2.re Re-estimated total sample size of group 2, before the bounds; with `ss.method = 'exact'` at most the largest size that `N.max` allows

N.re Re-estimated total sample size

n1.stage2 Number of additional patients to enrol in group 1

n2.stage2 Number of additional patients to enrol in group 2

n.stage2 Total number of additional patients to enrol

N1.final Final sample size of group 1

N2.final Final sample size of group 2

N.final Final total sample size

Power Exact power at the final sample size under the recovered proportions

The attribute `search` holds the search, and the attribute `search.limit` the limit of the size of group 2 examined by `search = 'stable'` (otherwise `NA`).

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

`BinaryPowerBSSR`, `BinarySampleSize`

Examples

```
# Interim data: 20 patients per group, 11 responders in total
BinaryBSSR(n1 = 20, n2 = 20, S = 11, Delta.A = 0.3, r = 1,
           alpha = 0.025, tar.power = 0.8, Test = 'Chisq')


# Restricted rule with a planned sample size of 40 per group
BinaryBSSR(n1 = 20, n2 = 20, S = 11, Delta.A = 0.3, r = 1,
           alpha = 0.025, tar.power = 0.8, Test = 'Boschloo',
           restricted = TRUE, N1 = 40, N2 = 40)
```

BinaryCondRejectBSSR *Conditional Rejection Probabilities of a Blinded Sample Size Re-estimation Design*

Description

Computes, under the null hypothesis of equal response probabilities, the probability that a two-arm trial with a binary endpoint and blinded sample size re-estimation (BSSR) rejects the null hypothesis given the pooled number of responders at the interim analysis and in the second stage. These conditional rejection probabilities do not depend on the common response probability, and the type I error rate is their average over the binomial distribution of the two pooled counts. With `theta`, the type I error rate is also returned for each pooled number of interim responders.

Usage

```
BinaryCondRejectBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard", "null.variance", "alternative.variance"),
  ss.Test = Test,
  ss.alpha = alpha,
  rounding = c("group", "friede-kieser", "total", "nearest"),
```

```
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  theta = NULL,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, used to split the blinded pooled proportion into group-specific proportions
N1	Initial sample size of group 1
N2	Initial sample size of group 2
omega	Fraction of the initial sample size observed at the interim analysis. The interim size of group 2 is ceiling(omega N2) and that of group 1 is ceiling(r ceiling(omega N2)), so the interim analysis keeps the allocation ratio. Either omega or n.interim must be supplied
r	Allocation ratio to group 1
alpha	Level of significance of the final analysis, for the alternative specified by alternative
tar.power	Target power
Test	Type of statistical test of the final analysis. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
restricted	Logical. If TRUE, the re-estimated sample size is not allowed to fall below the initial sample size. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), 'less', 'two.sided', or 'equivalence' for the alternative that the effect lies between the margins of a test of equivalence, see margin
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference p1 - p2, 'RR' for the risk ratio p1 / p2 or 'OR' for the odds ratio
ss.method	How the sample size is re-estimated from the recovered proportions. Options: 'exact' (default), 'standard', 'null.variance' or 'alternative.variance', as in BinarySampleSize
ss.Test	Test whose exact power is used for the re-estimation when ss.method = 'exact'. Default is Test
ss.alpha	Level of significance used for the re-estimation. Default is alpha

rounding	How the re-estimated sample size is turned into whole numbers. Options: 'group' (default), 'friede-kieser', 'total' or 'nearest'. Only 'group' is available with ss.method = 'exact'. See Details
search	How the exact re-estimation chooses the sample size when ss.method = 'exact': 'crossing' (default), 'smallest' or 'stable', as in BinarySampleSize, without going beyond the largest size that N.max allows. Ignored by the other methods
search.limit	Limit of the size of group 2 examined by search = 'stable', as in BinarySampleSize. Default is c(2, 50)
N.min	Lower bound on the final total sample size, or NULL (default) for none beyond the interim total. It can be used to keep the patients who are already enrolled but not yet evaluated at the interim analysis
N.max	Upper bound on the final total sample size, or NULL (default) for none
n.interim	Interim sample sizes of group 1 and group 2, as a vector of length two. An alternative to omega
theta	Optional vector of common response probabilities at which the type I error rate is decomposed by the pooled number of interim responders. Default is NULL, which skips the decomposition
margin	Must be 0, the default, since only tests of superiority are covered (see Details). The argument keeps the arguments the same as those of BinaryTypeIErrorBSSR
ref.pvalue	Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is FALSE. It applies to the final analysis and to the exact re-estimation

Details

Let n11 and n12 be the interim sample sizes of the two groups, and let n21(s) and n22(s) be the second-stage sizes reached from the pooled number s of interim responders. Under the null hypothesis both groups share the response probability theta. Given s, the number of interim responders in group 1 is hypergeometric, being the number of group 1 patients among s responders drawn from n11 + n12 patients. Given the pooled number s2 of second-stage responders, the number of second-stage responders in group 1 is an independent hypergeometric count of the same kind. CRP sums the rejection region of the final sample size over these two distributions and does not depend on theta. The type I error rate is sum_s sum_s2 b(s; n11 + n12, theta) b(s2; n21(s) + n22(s), theta) CRP(s, s2), where b is the binomial probability.

CRP.total is the conditional size of the final test given the total number of responders, with the responder count of group 1 hypergeometric among all final patients.

Only tests of superiority are covered. With a non-inferiority margin the two response probabilities differ on the boundary of the null hypothesis, and the conditional distributions of the responder counts then depend on them.

Value

An object of class bbssr_crp, a data frame with one row for every pair of pooled responder counts that the design can produce, ordered by s and then by s2, containing:

s Pooled number of responders at the interim analysis

s2 Pooled number of responders in the second stage

N1 Final sample size of group 1 reached from `s`

N2 Final sample size of group 2 reached from `s`

CRP Rejection probability given `s` and `s2`

CRP.total Rejection probability given only the total number of responders `s + s2`, when the responder count of group 1 is treated as one hypergeometric count among all `N1 + N2` patients

The attribute `max` is a data frame with the largest value of each of the two probabilities and the first outcome, in the order of the rows, at which it is attained up to rounding error, the attribute `exceed` gives the number of outcomes at which each of them exceeds `alpha`, and the attribute `reestimation` holds the final sample size for every pooled number of interim responders. When `theta` is supplied, the attribute `by.s` is a data frame with one row for each value of `theta` and of `s`, containing the probability `prob.s` of `s`, the type I error rate `TIE.s` given `s` and the contribution `prob.s * TIE.s` of `s` to the type I error rate, and the attribute `TIE` gives the type I error rate at each value of `theta`.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

BinaryTypeIErrorBSSR, BinaryPowerBSSR

Examples

```
crp <- BinaryCondRejectBSSR(
  Delta.A = 0.3, N1 = 12, N2 = 12, n.interim = c(6, 6), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = c(0.3, 0.5)
)
print(crp)
attr(crp, 'TIE')


crp <- BinaryCondRejectBSSR(
  Delta.A = 0.3, N1 = 39, N2 = 39, n.interim = c(20, 20), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard'
)
attr(crp, 'max')
plot(crp)
```

BinaryGridBSSR *Power and Sample Size of Several Blinded Sample Size Re-estimation Designs*

Description

Evaluates a set of designs with blinded sample size re-estimation (BSSR) for a binary endpoint over a common set of true pooled response probabilities, by calling BinaryPowerBSSR for each design, and returns the results as one data frame with a row for each design and each probability.

Usage

```
BinaryGridBSSR(design, p, Delta.T = NULL, type1 = FALSE, verbose = FALSE, ...)
```

Arguments

design	Data frame with one row per design. Its columns are arguments of BinaryPowerBSSR other than p and n.interim, for example Test, Delta.A, omega or r. The interim sample sizes are given by the two columns n1.interim and n2.interim, and the initial sample sizes either by the columns N1 and N2 or by a column p.plan, see Details. A missing value in a column means that the argument is not given for that design, so that designs specified by omega and by the interim sample sizes can share one data frame
p	Vector of true pooled response probabilities at which every design is evaluated. Values at which a response probability falls outside the unit interval are dropped for that design, and every design needs at least one value that is kept
Delta.T	True treatment effect. The default of NULL uses the assumed effect Delta.A of each design. It can also be given as a column of design, but not in both places
type1	Logical. If TRUE, the largest type I error rate of each design and of the corresponding fixed-sample design is added, as computed by BinaryTypeIErrorBSSR with its default grid of theta and the certified maximum. It is taken over the common response probability, over the boundary of the null hypothesis for a design with a non-inferiority margin, or over both boundaries for a test of equivalence. Default is FALSE
verbose	Logical. If TRUE, a message is issued as each design is evaluated. Default is FALSE
...	Arguments shared by all designs, with the same names as the columns of design. An argument may not be given both here and as a column

Details

When design has a column p.plan instead of the initial sample sizes, the planning proportions of the two groups are obtained from the pooled proportion p.plan and the assumed effect Delta.A, as in the re-estimation, and the initial sample sizes are those of BinarySampleSize with the test, the level, the target power and the allocation ratio of the design, the method ss.method of the

re-estimation and the rounding rule of the design. The test `Test` and the level `alpha` of the final analysis are used even when the re-estimation uses another test `ss.Test` or level `ss.alpha`.

The quartiles and `P.N.max` are those of `summary.bbssr_powerbssr`. The p-values of each test are kept for the rest of the session, so designs that share sample sizes are evaluated faster after the first one.

Value

An object of class `bbssr_grid`, a data frame with one row for each design and each element of `p` that gives response probabilities inside the unit interval, containing the index `design` of the design, the columns of `design` (other than the sample sizes and `Delta.T`), and:

N1, N2 Initial sample sizes

n1.interim, n2.interim Interim sample sizes

Delta.T True treatment effect

p, p1, p2 True pooled response probability and those of the two groups

power.BSSR, power.TRAD Power of the BSSR design and of the fixed-sample design with the initial sample sizes

E.N, SD.N Expected value and standard deviation of the final total sample size

N.q25, N.q50, N.q75 Quartiles of the final total sample size

P.N.max Probability that the final total sample size is the largest the design can reach

The attribute `designs` has one row per design with its index, the columns of `design` other than the sample sizes and `Delta.T`, the initial and interim sample sizes and the true effect and, with `type1 = TRUE`, the largest type I error rates `TIE.BSSR` and `TIE.TRAD` and the values `theta.BSSR` and `theta.TRAD` at which they occur. When a design tests equivalence, the columns `boundary.BSSR` and `boundary.TRAD` give the boundary, `'lower'` or `'upper'`, of each largest rate, and are `NA` for the other designs. Warnings and errors raised while a design is evaluated are prefixed with the index of the design.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

`BinaryPowerBSSR`, `BinaryTypeIErrorBSSR`, `BinarySampleSize`

Examples

```
design <- expand.grid(Test = c('Chisq', 'Fisher'), omega = c(0.3, 0.5),
                      stringsAsFactors = FALSE)
grid <- BinaryGridBSSR(design, p = c(0.3, 0.4, 0.5), Delta.A = 0.3, N1 = 30, N2 = 30,
                       r = 1, alpha = 0.025, tar.power = 0.8, ss.method = 'standard')
summary(grid)


plot(grid, colour.by = 'Test', facet.by = 'omega')
```

```
# Initial sample sizes from a planning proportion, and the largest type I error rates
design <- data.frame(Delta.A = c(0.2, 0.3), p.plan = 0.4)
grid <- BinaryGridBSSR(design, p = c(0.3, 0.5), omega = 0.5, r = 1, alpha = 0.025,
                       tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
                       type1 = TRUE)
attr(grid, 'designs')
```

`BinaryPower` *Power Calculation for Two-Arm Trials with Binary Endpoints*

Description

Calculates the exact power for two-arm trials with binary endpoints. Seven tests are supported (see `BinaryRR`), each of which can be applied with a one-sided or a two-sided alternative, and two of them also test non-inferiority with a margin on the scale of the risk difference or the risk ratio. Vectors of response probabilities are accepted.

Usage

```
BinaryPower(
  p1,
  p2,
  N1,
  N2,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

`p1`	True probability of responders for group 1 (can be a vector)
`p2`	True probability of responders for group 2 (can be a vector of the same length as `p1`)
`N1`	Sample size for group 1
`N2`	Sample size for group 2
`alpha`	Level of significance for the alternative specified by `alternative`

Test	Type of statistical test. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), 'less', 'two.sided', or 'equivalence' for the alternative that the effect lies between the margins of a test of equivalence, see margin
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
margin	Non-inferiority margin, on the scale given by margin.scale. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p1 - p2 <= -margin against p1 - p2 > -margin when alternative is 'greater', and p1 - p2 >= margin against p1 - p2 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio R0, and the null hypothesis is p1 / p2 <= R0 against p1 / p2 > R0 when alternative is 'greater', and p1 / p2 >= R0 against p1 / p2 < R0 when it is 'less'. A margin other than 0, and any margin on the scale of the risk ratio, requires Test = 'Blackwelder' or 'Farrington-Manning' and a one-sided alternative or 'equivalence', see BinaryRR. For alternative = 'equivalence' the margin is the pair c(lower, upper) of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis p1 - p2 <= -lower or p1 - p2 >= upper is tested against -lower < p1 - p2 < upper. On the scale of the risk ratio the pair satisfies 0 < lower < 1 < upper, a single ratio R gives the pair of R and 1 / R, the smaller as lower, and the null hypothesis is p1 / p2 <= lower or p1 / p2 >= upper. The null hypothesis is rejected when both one-sided tests reject, each at the level alpha
ref.pvalue	Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is FALSE
margin.scale	Scale of margin. Options: 'RD' (default) for the risk difference p1 - p2 or 'RR' for the risk ratio p1 / p2

Details

The power is obtained by summing the joint probability mass function of the two independent binomial counts over the rejection region returned by BinaryRR. The summation covers the whole rejection region rather than a row-wise tail, so it remains valid for two-sided tests, whose rejection regions are not contiguous within a row of the outcome grid.

Value

An object of class bbssr_power, a data frame with one row per element of p1 containing:

p1 True probability of responders for group 1

p2 True probability of responders for group 2

N1 Sample size for group 1

N2 Sample size for group 2

alpha Level of significance

Test Name of the statistical test

alternative Direction of the alternative hypothesis

Power Exact power

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

Examples

```
# Power of the one-sided chi-squared test
BinaryPower(p1 = 0.5, p2 = 0.2, N1 = 5, N2 = 5, alpha = 0.025, Test = 'Chisq')


# Power over a range of response probabilities for the two-sided Boschloo test
pw <- BinaryPower(p1 = c(0.5, 0.6, 0.7, 0.8), p2 = rep(0.2, 4),
                  N1 = 20, N2 = 20, alpha = 0.05, Test = 'Boschloo',
                  alternative = 'two.sided')
print(pw)
plot(pw)
```

`BinaryPowerBSSR` *Power of a Blinded Sample Size Re-estimation Design for Binary Endpoints*

Description

Calculates the power of a two-arm trial with a binary endpoint when blinded sample size re-estimation (BSSR) is implemented, together with the power of the corresponding fixed-sample design, the expected final sample size and its distribution. Seven tests are supported (see `BinaryRR`), each of which can be applied with a one-sided or a two-sided alternative, under either a restricted or an unrestricted design rule.

Usage

```
BinaryPowerBSSR(
  p,
  Delta.A,
  Delta.T,
  N1,
```

```
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard", "null.variance", "alternative.variance"),
  ss.Test = Test,
  ss.alpha = alpha,
  rounding = c("group", "friede-kieser", "total", "nearest"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

p	Vector of true pooled proportions of responders from both groups
Delta.A	Assumed treatment effect, on the scale given by effect, used to split the blinded pooled proportion into group-specific proportions
Delta.T	True treatment effect, on the scale given by effect
N1	Initial sample size of group 1
N2	Initial sample size of group 2
omega	Fraction of the initial sample size observed at the interim analysis. The interim size of group 2 is ceiling(omega N2) and that of group 1 is ceiling(r ceiling(omega N2)), so the interim analysis keeps the allocation ratio. Either omega or n.interim must be supplied
r	Allocation ratio to group 1
alpha	Level of significance of the final analysis, for the alternative specified by alternative
tar.power	Target power
Test	Type of statistical test of the final analysis. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
restricted	Logical. If TRUE, the re-estimated sample size is not allowed to fall below the initial sample size. Default is FALSE

`alternative` Direction of the alternative hypothesis. Options: `'greater'` (default), `'less'`, `'two.sided'`, or `'equivalence'` for the alternative that the effect lies between the margins of a test of equivalence, see `margin`

`tsmethod` Convention used to construct the two-sided version of the conditional tests, see `BinaryRR`. Options: `'minlike'` (default), `'central'` or `'blaker'`

`n.grid` Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100

`bb.gamma` Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure

`effect` Scale of `Delta.A` and `Delta.T`. Options: `'RD'` (default) for the risk difference `p1 - p2`, `'RR'` for the risk ratio `p1 / p2` or `'OR'` for the odds ratio

`ss.method` How the sample size is re-estimated from the recovered proportions. Options: `'exact'` (default), `'standard'`, `'null.variance'` or `'alternative.variance'`, as in `BinarySampleSize`

`ss.Test` Test whose exact power is used for the re-estimation when `ss.method = 'exact'`. Default is `Test`

`ss.alpha` Level of significance used for the re-estimation. Default is `alpha`

`rounding` How the re-estimated sample size is turned into whole numbers. Options: `'group'` (default), `'friede-kieser'`, `'total'` or `'nearest'`. Only `'group'` is available with `ss.method = 'exact'`. See Details

`search` How the exact re-estimation chooses the sample size when `ss.method = 'exact'`: `'crossing'` (default), `'smallest'` or `'stable'`, as in `BinarySampleSize`, without going beyond the largest size that `N.max` allows. Ignored by the other methods

`search.limit` Limit of the size of group 2 examined by `search = 'stable'`, as in `BinarySampleSize`. Default is `c(2, 50)`

`N.min` Lower bound on the final total sample size, or `NULL` (default) for none beyond the interim total. It can be used to keep the patients who are already enrolled but not yet evaluated at the interim analysis

`N.max` Upper bound on the final total sample size, or `NULL` (default) for none

`n.interim` Interim sample sizes of group 1 and group 2, as a vector of length two. An alternative to `omega`

`margin` Non-inferiority margin, on the scale given by `margin.scale`. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis `p1 - p2 <= -margin` against `p1 - p2 > -margin` when `alternative` is `'greater'`, and `p1 - p2 >= margin` against `p1 - p2 < margin` when it is `'less'`. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio `R0`, and the null hypothesis is `p1 / p2 <= R0` against `p1 / p2 > R0` when `alternative` is `'greater'`, and `p1 / p2 >= R0` against `p1 / p2 < R0` when it is `'less'`. A margin other than 0, and any margin on the scale of the risk ratio, requires `Test = 'Blackwelder'` or `'Farrington-Manning'` and a one-sided alternative or `'equivalence'`, see `BinaryRR`. With a margin other than 0 on the scale of the risk difference the assumed and the true effects are risk differences (`effect =`

'RD'), and with a margin on the scale of the risk ratio they are risk ratios (effect = 'RR'). For alternative = 'equivalence' the margin is the pair c(lower, upper) of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis p1 - p2 <= -lower or p1 - p2 >= upper is tested against -lower < p1 - p2 < upper. On the scale of the risk ratio the pair satisfies 0 < lower < 1 < upper, a single ratio R gives the pair of R and 1 / R, the smaller as lower, and the null hypothesis is p1 / p2 <= lower or p1 / p2 >= upper. The null hypothesis is rejected when both one-sided tests reject, each at the level alpha

ref.pvalue — Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is FALSE. It applies to the final analysis, to the fixed-sample comparator and to the exact re-estimation

margin.scale — Scale of margin. Options: 'RD' (default) for the risk difference p1 - p2 or 'RR' for the risk ratio p1 / p2

Details

At the interim analysis the pooled number of responders is observed without unblinding. The pooled proportion is combined with the assumed treatment effect Delta.A to recover group-specific proportions, from which the sample size is re-estimated. For the risk difference the recovered proportions are hat.p + Delta.A / (1 + r) and hat.p - r Delta.A / (1 + r). For the risk ratio they are the proportions with the pooled value hat.p and the ratio Delta.A, formula (21.11) of Kieser (2020), and for the odds ratio they are obtained from the quadratic equation that the odds ratio and the pooled value define. The exact search of ss.method = 'exact' uses the recovered proportions truncated to the unit interval. The normal approximation uses them before truncation, with each Bernoulli variance truncated at zero, so that the assumed effect is kept as in formula (2) of Friede and Kieser (2004). The power is then averaged over the distribution of the interim outcome.

Under rounding = 'group', both the interim and the final sample size of group 1 are obtained from the size of group 2 by a single application of ceiling(r ...), so the allocation ratio is preserved as closely as whole numbers allow and is exact whenever r is a whole number. Under rounding = 'friede-kieser', which requires a normal approximation, that is an ss.method other than 'exact', the second stage receives the rounded-up excess of the unrounded re-estimated total over the interim total, split into ceiling(n / (1 + r)) and ceiling(r n / (1 + r)) patients as in Friede and Kieser (2004). Under rounding = 'total' the unrounded total is rounded up and group 2 receives floor(N / (1 + r)) patients. Under rounding = 'nearest' the unrounded total is split in the ratio r to 1 and each group is rounded to the nearest whole number, which reproduces the computations of Friede et al. (2007). In each case the final total is kept between the interim total (or N.min, or the initial total under the restricted rule) and N.max, except that 'friede-kieser' and 'nearest' round the two groups separately and can exceed N.max by one patient. The argument N1 enters only through the fixed-sample comparator and the restricted rule, and under rounding = 'group' a warning is issued when it is not ceiling(r N2).

Recovered proportions that coincide admit no sample size. The risk ratio produces them when no interim patient responds, and the odds ratio when no interim patient or every interim patient responds. The initial sample size is kept for such interim outcomes.

Setting Delta.T to the null value (0 for the risk difference, 1 for a ratio) makes the two groups identical, so power.BSSR and power.TRAD become rejection probabilities under the null hypothesis. BinaryTypeIErrorBSSR evaluates them over the unit interval and locates the maximum.

With a non-inferiority margin, the design of Friede et al. (2007) is obtained. The assumed effect Delta.A is usually 0, which recovers the pooled proportion for both groups, the sample size is re-estimated by the formula of Farrington and Manning (ss.method = 'standard') or of Blackwelder (ss.method = 'alternative.variance'), and Delta.T = -margin (or margin for alternative = 'less') gives the rejection probability on the boundary of the null hypothesis.

With margin.scale = 'RR' the margin is a ratio R0 and the effects are risk ratios (effect = 'RR'). The assumed effect Delta.A must exceed R0 for alternative = 'greater' and fall below it for 'less', and is usually 1. The sample size is re-estimated by formula (8) of Farrington and Manning (1990) (ss.method = 'standard') or by their Method 1 (ss.method = 'alternative.variance'), and Delta.T = R0 gives the rejection probability on the boundary of the null hypothesis. The recovered proportions lie on that boundary only when no interim patient responds, and the initial sample size is kept for this interim outcome.

Value

An object of class bbssr_powerbssr, a data frame with one row per element of p containing:

p1 True probability of responders for group 1

p2 True probability of responders for group 2

p True pooled probability of responders from both groups

power.BSSR Power of the BSSR design

power.TRAD Power of the fixed-sample design

E.N Expected total sample size of the BSSR design

The interim sample sizes are stored as the attributes n1.interim and n2.interim. The attribute reestimation holds, for every pooled number of interim responders s, the recovered proportions, the final sample sizes and, in the column N2.limit, the limit of search = 'stable' (otherwise NA). The attribute N.dist holds the distribution of the final sample size for every row of the result, identified by the column scenario, and the attribute search the search. summary() reports the standard deviation and quantiles of the final sample size.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Friede T, Kieser M (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics*, 3(4), 269-279.

Kieser M (2020). *Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials*. Springer, Cham.

Friede T, Mitchell C, Mueller-Velten G (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal*, 49(6), 903-916.

See Also

BinaryTypeIErrorBSSR, BinaryAlphaAdjBSSR, BinaryBSSR

Examples

```
# Small BSSR calculation with the chi-squared test
BinaryPowerBSSR(
  p = 0.45,
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 5, N2 = 5, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq'
)


res <- BinaryPowerBSSR(
  p = seq(0.23, 0.47, by = 0.04),
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 39, N2 = 39, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Z-pool'
)
print(res)
summary(res)
plot(res)

# Re-estimation by the normal approximation with an upper bound of twice the
# initial sample size, and a final analysis with the Boschloo test
BinaryPowerBSSR(
  p = seq(0.2, 0.4, by = 0.05),
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 30, N2 = 30, n.interim = c(15, 15), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Boschloo',
  ss.method = 'standard', N.max = 120
)

# Non-inferiority with a margin of 0.1, as in Friede et al. (2007)
BinaryPowerBSSR(
  p = c(0.5, 0.7), Delta.A = 0, Delta.T = 0,
  N1 = 329, N2 = 329, n.interim = c(66, 66), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Farrington-Manning',
  ss.method = 'standard', rounding = 'nearest', margin = 0.1
)
```

`BinaryRR` *Rejection Region for Two-Arm Trials with Binary Endpoints*

Description

Provides a rejection region (RR) for two-arm trials with binary endpoints. Seven tests are supported. Each can be applied with a one-sided or a two-sided alternative, and two of them also test non-inferiority with a margin on the scale of the risk difference or the risk ratio.

Usage

```
BinaryRR(
  N1,
  N2,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

N1	Sample size for group 1
N2	Sample size for group 2
alpha	Level of significance for the alternative specified by alternative
Test	Type of statistical test. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for the one-sided alternative that the response probability of group 1 exceeds that of group 2, 'less' for the one-sided alternative that it falls below that of group 2, 'two.sided', or 'equivalence' for the alternative that the effect lies between the margins of a test of equivalence, see margin
tsmethod	Convention used to construct the two-sided version of the conditional tests, see Details. Options: 'minlike' (default), 'central' or 'blaker'. The Boschloo test orders the outcomes by the two-sided Fisher p-value of the selected convention. Ignored for a one-sided alternative, and ignored by 'Chisq', 'Z-pool', 'Blackwelder' and 'Farrington-Manning', whose two-sided versions are based on the absolute value of the Z statistic
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100. Ignored by the conditional tests
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the unconditional tests. The default of 0 disables the procedure. A positive value restricts the search over the nuisance parameter to an exact 100 (1 - bb.gamma) percent confidence interval and adds bb.gamma to the resulting p-value. A common choice is 0.0001
margin	Non-inferiority margin, on the scale given by margin.scale. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p1 - p2 <= -margin against p1 - p2 > -margin when alternative is 'greater', and p1 - p2 >= margin against p1 - p2 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio R0, and

the null hypothesis is p1 / p2 <= R0 against p1 / p2 > R0 when alternative is 'greater', and p1 / p2 >= R0 against p1 / p2 < R0 when it is 'less'. A margin other than 0, and any margin on the scale of the risk ratio, requires Test = 'Blackwelder' or 'Farrington-Manning' and a one-sided alternative or 'equivalence', see BinaryRR. For alternative = 'equivalence' the margin is the pair c(lower, upper) of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis p1 - p2 <= -lower or p1 - p2 >= upper is tested against -lower < p1 - p2 < upper. On the scale of the risk ratio the pair satisfies 0 < lower < 1 < upper, a single ratio R gives the pair of R and 1 / R, the smaller as lower, and the null hypothesis is p1 / p2 <= lower or p1 / p2 >= upper. The null hypothesis is rejected when both one-sided tests reject, each at the level alpha

ref.pvalue Logical. If TRUE, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see Details. Default is FALSE. Ignored by the conditional tests

margin.scale Scale of margin. Options: 'RD' (default) for the risk difference p1 - p2 or 'RR' for the risk ratio p1 / p2

Details

The function supports the following seven tests:

- The Pearson chi-squared test (Chisq)
- The Fisher exact test (Fisher)
- The Fisher mid-p test (Fisher-midP)
- The Z-pooled exact unconditional test (Z-pool)
- The exact unconditional test of Boschloo (1970), which orders the outcomes by the Fisher p-value (Boschloo)
- The test of Blackwelder (1982) with the unpooled standard error (Blackwelder)
- The test of Farrington and Manning (1990) with the standard error at the restricted maximum likelihood estimates (Farrington-Manning)

For the two-sided versions of the conditional tests, 'minlike' sums the null probabilities of all tables that are no more likely than the observed table, which is the convention of stats::fisher.test, and 'central' doubles the smaller of the two one-sided tail probabilities. 'blaker' orders the tables by the smaller of their two one-sided tail probabilities and sums the null probabilities of all tables at which this is no larger than at the observed table, formula (2) of Mehrotra, Chan and Berger (2003). For the exact Fisher p-value this is the convention 'blaker' of the **exact2x2** package. The exact Fisher p-value of 'blaker' never exceeds that of 'central', but its mid-p value and its Boschloo p-value can. When the two groups are of equal size the conditional distribution is symmetric, so the three conventions give the same exact Fisher p-value and hence the same Boschloo p-value, and 'minlike' and 'blaker' also give the same mid-p value. Under the mid-p correction the tables tied with the observed table in the ordering of 'minlike' or 'blaker', the observed table included, contribute half of their probability, following the definition of the mid-p value in Fay and Hunsberger (2021, Section 9). The two-sided versions of 'Chisq' and 'Z-pool' order the outcomes by the absolute value of the Z statistic.

The unconditional tests maximize the null tail probability of an ordering statistic over the common response probability, which is a nuisance parameter. Outcomes sharing the same value of the ordering statistic receive the same p-value. With a positive `bb.gamma` the maximization is restricted to a confidence interval for the nuisance parameter and `bb.gamma` is added, following Berger and Boos (1994).

The tail probability of each outcome is a polynomial in the common response probability, and its maximum over the grid of `n.grid` points is a lower bound of the maximum over the unit interval. A p-value computed on the grid can therefore fall below the exact p-value, and the test can then exceed its nominal level. With `ref.pvalue = TRUE` the grid is extended by points equally spaced on the arcsine square-root scale, which resolve the narrow local maxima near 0 and 1, and every local maximum on the extended grid is refined by a safeguarded Newton iteration between its two neighbouring grid points. The grid maximum is kept when it is larger, so a refined p-value is never smaller than the p-value on the grid. In the designs examined during development the refined p-values agreed with a certified maximum over the unit interval to within 1e-12, but the refinement is a local search and does not guarantee the maximum. It takes longer than the grid alone, by roughly an order of magnitude with 500 patients per group.

The alternative `'less'` is handled by exchanging the two groups, testing the alternative `'greater'` and exchanging them back. Every test considered treats the groups symmetrically apart from the direction of the alternative, so this gives the lower-tail version of each test.

The tests of Blackwelder and of Farrington and Manning refer the statistic `z = (hat.p1 - hat.p2 + margin) / SE` to the standard normal distribution, which tests the null hypothesis `p1 - p2 <= -margin` for `alternative = 'greater'`. The standard error of the Blackwelder test uses the observed proportions. That of the Farrington-Manning test uses the maximum likelihood estimates under `p1 - p2 = -margin`, formula (12) of Farrington and Manning (1990). With `margin = 0` the Farrington-Manning test coincides with `'Chisq'` and the Blackwelder test is the Wald test. A standard error of 0, which arises only when both observed proportions are 0 or 1, gives `z = Inf` or `-Inf` according to the sign of the numerator, and `z = 0` when the numerator is 0 as well. These tests are asymptotic, and their rejection regions and power are computed exactly, as for the other tests. A margin other than 0 is available only with these two tests and a one-sided alternative.

With `margin.scale = 'RR'` the margin is a ratio `R0`, and both tests refer the statistic `z = (hat.p1 - R0 hat.p2) / SE` of Farrington and Manning (1990) to the standard normal distribution, which tests the null hypothesis `p1 / p2 <= R0` for `alternative = 'greater'`. The squared standard error is `p1 (1 - p1) / N1 + R0^2 p2 (1 - p2) / N2`. The Blackwelder test evaluates it at the observed proportions, which is Method 1 of Farrington and Manning (1990), and the Farrington-Manning test at the maximum likelihood estimates under `p1 = R0 p2`, formula (13) of that article. For `alternative = 'less'` the groups are exchanged as for the other tests, with the ratio `1 / R0`, which changes only the sign of the statistic. A margin on the scale of the risk ratio is also available only with these two tests and a one-sided alternative.

With `alternative = 'equivalence'` the null hypothesis `p1 - p2 <= -lower` or `p1 - p2 >= upper` (`p1 / p2 <= lower` or `p1 / p2 >= upper` on the scale of the risk ratio) is tested by two one-sided tests of the same kind, the test against `'greater'` with the margin `lower` and the test against `'less'` with the margin `upper`, each at the level `alpha`. The null hypothesis is rejected when both reject, so the rejection region is the intersection of the two one-sided regions and the p-value of an outcome is the larger of its two one-sided p-values. Tests of equivalence are available with the tests of Blackwelder and of Farrington and Manning.

The p-values are kept for the rest of the session and reused by later calls with the same sample sizes and test, see `bbssr-package`. They form a matrix of `(N1 + 1) (N2 + 1)` values, which takes about 2

GB of memory for 16,000 patients per group, so the memory and the time of the exact calculations grow with the product of the group sizes.

Value

An object of class `bbssr_rr`, which is a logical matrix of dimension `(N1 + 1)` by `(N2 + 1)` whose entry `[i + 1, j + 1]` is `TRUE` when the null hypothesis is rejected at `i` responders in group 1 and `j` responders in group 2. The design settings are stored as attributes

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Berger RL, Boos DD (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89(427), 1012-1016.

Blackwelder WC (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*, 3(4), 345-353.

Boschloo RD (1970). Raised conditional level of significance for the 2 x 2-table when testing the equality of two probabilities. *Statistica Neerlandica*, 24(1), 1-9.

Farrington CP, Manning G (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9(12), 1447-1454.

Fay MP, Hunsberger SA (2021). Practical valid inferences for the two-sample binomial problem. *Statistics Surveys*, 15, 72-110.

Mehrotra DV, Chan ISF, Berger RL (2003). A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. *Biometrics*, 59(2), 441-450.

Examples

```
# Simple example with small sample sizes
RR <- BinaryRR(N1 = 5, N2 = 5, alpha = 0.025, Test = 'Chisq')
print(RR)


# Two-sided Boschloo test with the Berger-Boos procedure
RR <- BinaryRR(N1 = 20, N2 = 10, alpha = 0.05, Test = 'Boschloo',
               alternative = 'two.sided', bb.gamma = 0.0001)
print(RR)
plot(RR)

# The grid maximum can understate the p-value. Refinement removes two outcomes from the
# rejection region of the Z-pooled test with 32 patients per group
RR.grid <- BinaryRR(N1 = 32, N2 = 32, alpha = 0.025, Test = 'Z-pool')
RR.ref <- BinaryRR(N1 = 32, N2 = 32, alpha = 0.025, Test = 'Z-pool', ref.pvalue = TRUE)
sum(RR.grid) - sum(RR.ref)

# Non-inferiority with a margin of 0.1 by the test of Farrington and Manning
```

```
RR <- BinaryRR(N1 = 60, N2 = 60, alpha = 0.025, Test = 'Farrington-Manning',
               margin = 0.1)
print(RR)

# Non-inferiority with the margin 0.8 on the scale of the risk ratio
RR <- BinaryRR(N1 = 60, N2 = 60, alpha = 0.025, Test = 'Farrington-Manning',
               margin = 0.8, margin.scale = 'RR')
print(RR)

# Equivalence with the margins -0.15 and 0.15 of the risk difference
RR <- BinaryRR(N1 = 100, N2 = 100, alpha = 0.05, Test = 'Farrington-Manning',
               alternative = 'equivalence', margin = c(0.15, 0.15))
print(RR)
```

`BinarySampleSize` *Sample Size Calculation for Two-Arm Trials with Binary Endpoints*

Description

Calculates the required sample size for two-arm trials with binary endpoints. Seven tests are supported (see `BinaryRR`), each of which can be applied with a one-sided or a two-sided alternative. The Blackwelder and Farrington-Manning tests also allow a non-inferiority margin and a test of equivalence. The sample size can be obtained from the exact power of the selected test or from the normal approximation.

Usage

```
BinarySampleSize(
  p1,
  p2,
  r,
  alpha,
  tar.power,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  method = c("exact", "standard", "null.variance", "alternative.variance"),
  rounding = c("group", "friede-kieser", "total", "nearest"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

p1	True probability of responders for group 1
p2	True probability of responders for group 2
r	Allocation ratio to group 1 (i.e., allocation ratio of group 1:group 2 = r:1, r > 0)
alpha	Level of significance for the alternative specified by alternative
tar.power	Target power
Test	Type of statistical test. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), which requires p1 - p2 > -margin, 'less', which requires p1 - p2 < margin, or 'two.sided', which requires p1 != p2. With margin.scale = 'RR', 'greater' requires p1 > margin p2 and 'less' requires p1 < margin p2. 'equivalence' requires -lower < p1 - p2 < upper, or lower < p1 / p2 < upper on the scale of the risk ratio
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
method	How the sample size is obtained. 'exact' (default) steps from the normal approximation, one patient at a time, to the sample size at which the exact power of Test first attains the target, see Details. 'standard' uses the normal approximation with the variance under the null hypothesis for the significance term and the variance under the alternative for the power term, formula (21.3) of Kieser (2020). 'null.variance' uses the variance under the null hypothesis for both terms, formula (1) of Friede and Kieser (2004). 'alternative.variance' uses the variance under the alternative for both terms, as in Blackwelder (1982). See Details for a non-zero margin and for a margin on the scale of the risk ratio
rounding	How an unrounded sample size from the normal approximation is turned into whole numbers. 'group' (default) rounds up the size of group 2 and gives group 1 ceiling(r N2) patients. 'friede-kieser' rounds up the two group sizes separately, as in Friede and Kieser (2004). 'total' rounds up the total and gives group 2 floor(N / (1 + r)) patients. 'nearest' rounds the two group sizes separately to the nearest whole number, as in Farrington and Manning (1990). Only 'group' is available with method = 'exact'
search	How the search of method = 'exact' chooses the size of group 2. 'crossing' (default) steps from the normal approximation to a size that attains the target power while the size one unit smaller does not. 'smallest' returns the smallest size that attains the target power. 'stable' returns the smallest size from which every size up to a limit attains it. See Details. Ignored by the other methods
search.limit	Two numbers a and b giving the limit max(ceiling(a n0), ceiling(n0 + b)) of the size of group 2 examined by search = 'stable', where n0 is the size of group 2 from the normal approximation, rounded up. a must be at least 1 and b non-negative. Default is c(2, 50)

`margin` Non-inferiority margin, on the scale given by `margin.scale`. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis `p1 - p2 <= -margin` against `p1 - p2 > -margin` when `alternative` is `'greater'`, and `p1 - p2 >= margin` against `p1 - p2 < margin` when it is `'less'`. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio `R0`, and the null hypothesis is `p1 / p2 <= R0` against `p1 / p2 > R0` when `alternative` is `'greater'`, and `p1 / p2 >= R0` against `p1 / p2 < R0` when it is `'less'`. A margin other than 0, and any margin on the scale of the risk ratio, requires `Test = 'Blackwelder'` or `'Farrington-Manning'` and a one-sided alternative or `'equivalence'`, see BinaryRR. For `alternative = 'equivalence'` the margin is the pair `c(lower, upper)` of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis `p1 - p2 <= -lower` or `p1 - p2 >= upper` is tested against `-lower < p1 - p2 < upper`. On the scale of the risk ratio the pair satisfies `0 < lower < 1 < upper`, a single ratio `R` gives the pair of `R` and `1 / R`, the smaller as `lower`, and the null hypothesis is `p1 / p2 <= lower` or `p1 / p2 >= upper`. The null hypothesis is rejected when both one-sided tests reject, each at the level `alpha`

`ref.pvalue` Logical. If `TRUE`, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is `FALSE`

`margin.scale` Scale of `margin`. Options: `'RD'` (default) for the risk difference `p1 - p2` or `'RR'` for the risk ratio `p1 / p2`

Details

With the default `search = 'crossing'` the calculation uses a three-step approach:

1. Calculate an initial sample size from the normal approximation to the chi-squared test
2. Evaluate the exact power at the initial sample size
3. If the exact power attains the target power, lower the size of group 2 one unit at a time as long as the power still attains it; otherwise raise it one unit at a time until the power attains it

The normal approximation of the first step uses `alpha` for a one-sided alternative, `alpha / 2` for a two-sided alternative, and `alpha` for each of the two one-sided tests of equivalence. Only the starting value of the search is affected, so the returned sample size is exact in every case.

The search stops where the exact power crosses the target: the returned size of group 2 attains the target power and the size one unit smaller does not, unless the returned size is 1. The exact power is not monotone in the sample size, so a smaller size further from the normal approximation may also attain the target power, and a larger size may fall short of it.

The two other searches do not depend on where the normal approximation starts. `search = 'smallest'` scans the size of group 2 upwards from 1 and returns the first size that attains the target power. It visits every smaller size, which takes longer for the unconditional tests. `search = 'stable'` scans downwards from the limit given by `search.limit` and returns the smallest size from which every size up to the limit attains the target power, so that adding patients within this range does not take the power below the target. The search stops with an error if the limit itself falls short of the target power. The three searches can return different sizes because the exact power is not monotone in the sample size.

Under method = 'standard', 'null.variance' or 'alternative.variance' the steps above are replaced by the closed-form normal approximation and the rounding rule selected by rounding. The Power column still reports the exact power of Test at the resulting sample size.

With a non-zero margin, the difference p1 - p2 in the normal approximation is replaced by its distance p1 - p2 + margin from the boundary of the null hypothesis, and the variance under the null hypothesis is evaluated at the large sample values of the restricted maximum likelihood estimates of Farrington and Manning (1990). The method 'standard' then gives formula (4) of Farrington and Manning (1990), which is formula (2) of Friede et al. (2007), and 'alternative.variance' gives the formula of Blackwelder (1982), which is formula (1) of Friede et al. (2007). The exact search starts from the formula of Farrington and Manning.

With margin.scale = 'RR' and the ratio R0 = margin, the difference in the normal approximation is the distance p1 - R0 p2, or R0 p2 - p1 for alternative = 'less', and the variance of the size of group 2 is p1 (1 - p1) / r + R0^2 p2 (1 - p2), evaluated at the large sample values of the restricted maximum likelihood estimates under p1 = R0 p2 for the significance term. The method 'standard' then gives formula (8) of Farrington and Manning (1990), 'alternative.variance' gives the formula of their Method 1, and the exact search starts from formula (8).

For alternative = 'equivalence' the approximate power of each one-sided test is obtained as above for its own margin, at the level alpha, and the sample size of the normal approximation is the smallest unrounded size of group 2 at which the sum of the two approximate powers minus 1 attains the target power. The sum minus 1 is a lower bound of the approximate probability that both tests reject. The exact search starts from this size.

Value

An object of class bbssr_samplesize, a data frame with one row containing:

p1 True probability of responders for group 1

p2 True probability of responders for group 2

r Allocation ratio to group 1

alpha Level of significance

tar.power Target power

Test Name of the statistical test

alternative Direction of the alternative hypothesis

Power Exact power of Test at the selected sample size

N1 Required sample size of group 1

N2 Required sample size of group 2

N Total required sample size

The attribute search holds the search, and the attribute search.limit the limit of the size of group 2 examined by search = 'stable' (otherwise NA).

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Friede T, Kieser M (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics*, 3(4), 269-279.

Kieser M (2020). *Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials*. Springer, Cham.

Blackwelder WC (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*, 3(4), 345-353.

Farrington CP, Manning G (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9(12), 1447-1454.

Friede T, Mitchell C, Mueller-Velten G (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal*, 49(6), 903-916.

Examples

```
# One-sided chi-squared test
BinarySampleSize(p1 = 0.4, p2 = 0.2, r = 1, alpha = 0.025,
                 tar.power = 0.8, Test = 'Chisq')


# Two-sided Fisher exact test
BinarySampleSize(p1 = 0.5, p2 = 0.2, r = 2, alpha = 0.05,
                 tar.power = 0.9, Test = 'Fisher', alternative = 'two.sided')

# Normal approximation with the variance under the null hypothesis, two-sided test
# at level 0.05, as in Friede and Kieser (2004)
BinarySampleSize(p1 = 0.4, p2 = 0.2, r = 1, alpha = 0.05, tar.power = 0.8,
                 Test = 'Chisq', alternative = 'two.sided',
                 method = 'null.variance')

# Non-inferiority with a margin of 0.1, formula (2) of Friede et al. (2007)
BinarySampleSize(p1 = 0.7, p2 = 0.7, r = 1, alpha = 0.025, tar.power = 0.8,
                 Test = 'Farrington-Manning', method = 'standard',
                 rounding = 'friede-kieser', margin = 0.1)

# Non-inferiority on the scale of the risk ratio: the null hypothesis p1 / p2 <= 0.8
# with equal response probabilities of 0.6
BinarySampleSize(p1 = 0.6, p2 = 0.6, r = 1, alpha = 0.025, tar.power = 0.8,
                 Test = 'Farrington-Manning', margin = 0.8, margin.scale = 'RR')
```

`BinaryTypeIErrorBSSR` *Type I Error Rate of a Blinded Sample Size Re-estimation Design*

Description

Evaluates the type I error rate of a two-arm trial with a binary endpoint and blinded sample size re-estimation (BSSR) over the common response probability, or over the pooled response probability on the boundary of a non-inferiority hypothesis, together with that of the corresponding fixed-sample design, and locates the largest value of each.

Usage

```
BinaryTypeIErrorBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  tsmethod = c("minlike", "central", "blaker"),
  n.grid = 100,
  bb.gamma = 0,
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard", "null.variance", "alternative.variance"),
  ss.Test = Test,
  ss.alpha = alpha,
  rounding = c("group", "friede-kieser", "total", "nearest"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  theta = seq(0, 1, by = 0.005),
  maximize = c("certified", "refined", "grid"),
  margin = 0,
  ref.pvalue = FALSE,
  margin.scale = c("RD", "RR")
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, used to split the blinded pooled proportion into group-specific proportions
N1	Initial sample size of group 1
N2	Initial sample size of group 2
omega	Fraction of the initial sample size observed at the interim analysis. The interim size of group 2 is ceiling(omega N2) and that of group 1 is ceiling(r

	ceiling(omega N2)), so the interim analysis keeps the allocation ratio. Either omega or n.interim must be supplied
r	Allocation ratio to group 1
alpha	Level of significance of the final analysis, for the alternative specified by alternative
tar.power	Target power
Test	Type of statistical test of the final analysis. Options: 'Chisq', 'Fisher', 'Fisher-midP', 'Z-pool', 'Boschloo', 'Blackwelder' or 'Farrington-Manning'
restricted	Logical. If TRUE, the re-estimated sample size is not allowed to fall below the initial sample size. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), 'less', 'two.sided', or 'equivalence' for the alternative that the effect lies between the margins of a test of equivalence, see margin
tsmethod	Convention used to construct the two-sided version of the conditional tests, see BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'
n.grid	Number of grid points used to search over the nuisance parameter of the unconditional tests. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure. The default of 0 disables the procedure
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference p1 - p2, 'RR' for the risk ratio p1 / p2 or 'OR' for the odds ratio
ss.method	How the sample size is re-estimated from the recovered proportions. Options: 'exact' (default), 'standard', 'null.variance' or 'alternative.variance', as in BinarySampleSize
ss.Test	Test whose exact power is used for the re-estimation when ss.method = 'exact'. Default is Test
ss.alpha	Level of significance used for the re-estimation. Default is alpha
rounding	How the re-estimated sample size is turned into whole numbers. Options: 'group' (default), 'friede-kieser', 'total' or 'nearest'. Only 'group' is available with ss.method = 'exact'. See Details
search	How the exact re-estimation chooses the sample size when ss.method = 'exact': 'crossing' (default), 'smallest' or 'stable', as in BinarySampleSize, without going beyond the largest size that N.max allows. Ignored by the other methods
search.limit	Limit of the size of group 2 examined by search = 'stable', as in BinarySampleSize. Default is c(2, 50)
N.min	Lower bound on the final total sample size, or NULL (default) for none beyond the interim total. It can be used to keep the patients who are already enrolled but not yet evaluated at the interim analysis
N.max	Upper bound on the final total sample size, or NULL (default) for none
n.interim	Interim sample sizes of group 1 and group 2, as a vector of length two. An alternative to omega

`theta` Grid of common response probabilities at which the type I error rate is evaluated. Default is `seq(0, 1, by = 0.005)`. With a non-zero `margin`, or a margin on the scale of the risk ratio, `theta` is the pooled response probability `(r p1 + p2) / (1 + r)` on the boundary of the null hypothesis, see Details

`maximize` How the largest type I error rate is located. `'certified'` (default) finds it over the whole interval from the smallest to the largest value of `theta`, restricted with a margin to the values at which both response probabilities on the null boundary lie in the unit interval, together with an upper bound, see Details. `'refined'` refines the largest local maxima on the grid by a one-dimensional optimization between the neighbouring grid points, and `'grid'` takes the largest value on the grid

`margin` Non-inferiority margin, on the scale given by `margin.scale`. On the scale of the risk difference the default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis `p1 - p2 <= -margin` against `p1 - p2 > -margin` when `alternative` is `'greater'`, and `p1 - p2 >= margin` against `p1 - p2 < margin` when it is `'less'`. A negative value tests for superiority by more than its absolute value. On the scale of the risk ratio the margin is a positive ratio `R0`, and the null hypothesis is `p1 / p2 <= R0` against `p1 / p2 > R0` when `alternative` is `'greater'`, and `p1 / p2 >= R0` against `p1 / p2 < R0` when it is `'less'`. A margin other than 0, and any margin on the scale of the risk ratio, requires `Test = 'Blackwelder'` or `'Farrington-Manning'` and a one-sided alternative or `'equivalence'`, see BinaryRR. With a margin other than 0 on the scale of the risk difference the assumed and the true effects are risk differences (`effect = 'RD'`), and with a margin on the scale of the risk ratio they are risk ratios (`effect = 'RR'`). For `alternative = 'equivalence'` the margin is the pair `c(lower, upper)` of margins of a test of equivalence, or a single value used for both. On the scale of the risk difference both margins lie in (0, 1), and the null hypothesis `p1 - p2 <= -lower` or `p1 - p2 >= upper` is tested against `-lower < p1 - p2 < upper`. On the scale of the risk ratio the pair satisfies `0 < lower < 1 < upper`, a single ratio `R` gives the pair of `R` and `1 / R`, the smaller as `lower`, and the null hypothesis is `p1 / p2 <= lower` or `p1 / p2 >= upper`. The null hypothesis is rejected when both one-sided tests reject, each at the level `alpha`

`ref.pvalue` Logical. If `TRUE`, the maximization over the nuisance parameter of the unconditional tests is refined between the grid points, see BinaryRR. Default is `FALSE`. It applies to the final analysis, to the fixed-sample comparator and to the exact re-estimation

`margin.scale` Scale of `margin`. Options: `'RD'` (default) for the risk difference `p1 - p2` or `'RR'` for the risk ratio `p1 / p2`

Details

Under the null hypothesis both groups share the response probability `theta`. The sample size is re-estimated exactly as in BinaryPowerBSSR, from the assumed effect `Delta.A`, and the rejection probability is summed over every interim outcome and every outcome of the second stage.

The type I error rate is a polynomial in `theta`, since both response probabilities are linear in `theta`. With `maximize = 'certified'` the polynomial is expressed in the Bernstein basis over the interval from the smallest to the largest value of `theta`, restricted with a margin to the values at which both

response probabilities lie in the unit interval. The basis polynomials are non-negative and sum to one, so on any subinterval the polynomial does not exceed its largest Bernstein coefficient there, and its first and last coefficients are its values at the two ends. The interval is halved repeatedly by the algorithm of de Casteljau, and a subinterval is set aside once its largest coefficient exceeds the largest value found by at most 1e-12. The largest coefficient of the subintervals set aside, reported as `bound`, is an upper bound of the type I error rate at every value of `theta` in the interval, up to rounding error, and `TIE` is within 1e-12 of it. If 1e5 subdivisions do not suffice, a warning is given; `bound` remains an upper bound, but `TIE` can then be more than 1e-12 below it. With `maximize = 'refined'` the three largest local maxima on the grid are refined, which usually finds the maximum when the grid is fine compared with the spacing of the local maxima but does not guarantee it, and with `maximize = 'grid'` the largest value on the grid is reported.

With a non-inferiority `margin` the null hypothesis is `p1 - p2 <= -margin`, or `p1 - p2 >= margin` for `alternative = 'less'`, and the type I error rate is evaluated on its boundary, as in Friede et al. (2007). The boundary is parametrized by the pooled response probability `theta`, and the values of `theta` at which a response probability falls outside the unit interval are dropped. With `margin.scale = 'RR'` the null hypothesis is `p1 / p2 <= margin`, or `p1 / p2 >= margin` for `alternative = 'less'`, and its boundary `p1 = margin p2` is parametrized in the same way, with `p2 = (1 + r) theta / (1 + r margin)`. Both response probabilities are again linear in `theta`, so the certified maximum applies on either scale.

Value

An object of class `bbssr_tie`, a data frame with one row per element of `theta` containing:

theta Common response probability of the two groups, or the pooled response probability on the null boundary

p1 Response probability of group 1

p2 Response probability of group 2

TIE.BSSR Type I error rate of the BSSR design

TIE.TRAD Type I error rate of the fixed-sample design with sample sizes `N1` and `N2`

The attribute `max` is a data frame with one row for each design, `'BSSR'` and `'TRAD'` in the column `Design`, holding the largest type I error rate `TIE`, the common response probability, or pooled response probability on the null boundary, `theta` at which it occurs and, with `maximize = 'certified'`, the upper bound `bound` of the type I error rate over the interval given by the attribute `interval` (otherwise `NA`). The attribute `reestimation` holds the final sample size for every pooled number of interim responders.

For `alternative = 'equivalence'` the null hypothesis has two boundaries, the lower one `p1 - p2 = -lower` (`p1 = lower p2` on the scale of the risk ratio) and the upper one `p1 - p2 = upper` (`p1 = upper p2`). The data frame then has a first column `boundary`, `'lower'` or `'upper'`, and a row for each element of `theta` on each boundary. The rejection probability is that of the two one-sided tests together. The largest rate of each design in the attribute `max` is the larger of its maxima on the two boundaries, its bound the larger of the two bounds, and the column `boundary` of the attribute gives the boundary on which it occurs. The attribute `interval` is then a matrix with a row for each boundary.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Friede T, Kieser M (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics*, 3(4), 269-279.

Kieser M (2020). *Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials*. Springer, Cham.

Friede T, Mitchell C, Mueller-Velten G (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal*, 49(6), 903-916.

See Also

BinaryAlphaAdjBSSR, BinaryPowerBSSR

Examples

```
tie <- BinaryTypeIErrorBSSR(
  Delta.A = 0.3, N1 = 20, N2 = 20, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = seq(0.05, 0.95, by = 0.05)
)
print(tie)


tie <- BinaryTypeIErrorBSSR(
  Delta.A = 0.3, N1 = 39, N2 = 39, n.interim = c(20, 20), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Boschloo'
)
attr(tie, 'max')
plot(tie)
```

PairedAlphaAdjBSSR *Adjusted Significance Level of Blinded Sample Size Re-Estimation for Paired Binary Data*

Description

Lowers the nominal significance level of a matched-pairs design with blinded sample size re-estimation until the largest type I error rate over the probability of a discordant pair on the boundary of the null hypothesis does not exceed the target level `alpha`, and returns the adjusted level together with that of the fixed design.

Usage

```
PairedAlphaAdjBSSR(
  Delta.A,
  N,
```

```
  omega = NULL,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  ss.method = c("exact", "standard", "conditional.power"),
  ss.Test = Test,
  ss.alpha = alpha,
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  psi = seq(0, 1, by = 0.01),
  maximize = c("certified", "refined", "grid"),
  adjust = c("test", "both"),
  tol = 1e-08,
  step = 1e-05,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

`Delta.A`	Assumed difference `p10 - p01` used in the re-estimation
`N`	Planned number of pairs, which is the number of pairs of the fixed design
`omega`	Fraction of `N` at the interim analysis, rounded up. Supply either `omega` or `n.interim`
`alpha`	Target level of significance, which the type I error rate must not exceed
`tar.power`	Target power of the re-estimation
`Test`	Type of statistical test. Options: `'McNemar'`, `'McNemar-cc'`, `'McNemar-exact'`, `'McNemar-midP'`, `'McNemar-uncond'` or `'Paired-Score'`
`restricted`	Logical. If `TRUE`, the final number of pairs is at least `N`. Default is `FALSE`
`alternative`	Direction of the alternative hypothesis. Options: `'greater'` (default) for `p10 - p01 > -margin`, `'less'` for `p10 - p01 < margin`, `'two.sided'` for `p10 != p01`, or `'equivalence'` for `-lower < p10 - p01 < upper`
`n.grid`	Number of grid points used to search over the probability of a discordant pair in the test `'McNemar-uncond'`. Default is 100
`bb.gamma`	Confidence level parameter of the Berger-Boos procedure for the test `'McNemar-uncond'`. The default of 0 disables the procedure
`ss.method`	Re-estimation rule. `'exact'` (default) and `'standard'` re-estimate the number of pairs of `PairedSampleSize` with the method of that name. `'conditional.power'` takes the smallest number of pairs at which the conditional power given the interim number of discordant pairs attains `tar.power`, the rule of Schepers et al. (2026); it requires `N.max`. See Details

ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level of significance used for the re-estimation when adjust = 'test'. Default is alpha. It is ignored when adjust = 'both'
search	Search of the exact re-estimation, see PairedSampleSize; it does not go beyond N.max
search.limit	Limit of the search 'stable', see PairedSampleSize
N.min	Smallest final number of pairs, or NULL
N.max	Largest final number of pairs, or NULL
n.interim	Number of pairs at the interim analysis. Supply either omega or n.interim
psi	Values of the probability p10 + p01 of a discordant pair at which the rate is evaluated. Values that do not lie on the boundary of the null hypothesis, below |margin|, are dropped. Default is seq(0, 1, by = 0.01)
maximize	How the largest type I error rate at a level is located, as in PairedTypeIErrorBSSR. With 'certified' (default) the adjusted level controls the type I error rate over the whole interval spanned by the values of psi on the boundary. With 'refined' and 'grid' the control is assessed on the grid
adjust	Which part of the design uses the adjusted level. 'test' (default) applies it to the final analysis only, so the re-estimation keeps using ss.alpha. 'both' applies it to the re-estimation as well
tol	Tolerance of the bisection, relative to alpha, a single value in (0, 1). Default is 1e-8
step	Decrement of the level in the search used when adjust = 'both'. Default is 1e-5
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

The searches are those of BinaryAlphaAdjBSSR. With adjust = 'test' the final numbers of pairs do not depend on the nominal level of the final analysis, the largest type I error rate is a non-decreasing function of the level, and the level is found by bisection to within tol * alpha; the reported level is the largest value with at most six significant digits that gives the same rejection regions. With adjust = 'both' the level is lowered from alpha in steps of step until the type I error rate is controlled. The fixed design always uses the bisection. With maximize = 'certified'

a level is accepted only when the upper bound of the type I error rate over the interval, computed as in PairedTypeIErrorBSSR, does not exceed alpha.

The exact conditional McNemar test keeps its level whatever blinded rule is used (Schepers et al., 2026), so it needs no adjustment.

Value

An object of class bbssr_alphaadj, a data frame with one row for the BSSR design and one for the fixed design containing:

Design 'BSSR' or 'TRAD'

alpha Target level

max.TIE Largest type I error rate at the nominal level alpha

alpha.adj Adjusted nominal level. It is alpha when the nominal level already controls the type I error rate, and 0, with a warning, when no level examined by the bisection does. The search of adjust = 'both' for the BSSR design stops with an error instead when it reaches 0

max.TIE.adj Largest type I error rate at the adjusted level

psi.adj Probability of a discordant pair at which max.TIE.adj occurs, NA when alpha.adj is 0

max.TIE.bound Upper bound of the type I error rate at the nominal level with maximize = 'certified', otherwise NA

max.TIE.adj.bound Upper bound of the type I error rate at the adjusted level with maximize = 'certified', otherwise NA

For alternative = 'equivalence' the column boundary.adj gives the boundary, 'lower' or 'upper', on which max.TIE.adj occurs.

References

Kieser M, Friede T (2000). Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine*, 19(7), 901-911.

Schepers M, Brannath W, Hoffmann E, Stingl J, Schmidtmann I (2026). Blinded sample size review for McNemar's test based on primary and surrogate endpoints. arXiv:2608.17784.

Examples

```
PairedAlphaAdjBSSR(Delta.A = 0.2, N = 40, n.interim = 20, alpha = 0.05,
                   tar.power = 0.8, Test = 'McNemar', alternative = 'two.sided',
                   ss.method = 'standard', psi = seq(0.05, 0.95, by = 0.05))
```

PairedBSSR *Blinded Sample Size Re-Estimation from Paired Interim Data*

Description

Re-estimates the number of pairs of a matched-pairs trial from the number of discordant pairs observed at an interim analysis, without the direction in which they are discordant.

Usage

```
PairedBSSR(
  n1,
  D1,
  Delta.A,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  N = NULL,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  ss.method = c("exact", "standard", "conditional.power"),
  ss.Test = Test,
  ss.alpha = alpha,
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

n1	Number of pairs at the interim analysis
D1	Number of discordant pairs among them
Delta.A	Assumed difference p10 - p01 used in the re-estimation
alpha	Level of significance for the alternative specified by alternative
tar.power	Target power of the re-estimation
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'
restricted	Logical. If TRUE, the final number of pairs is at least N. Default is FALSE
N	Planned number of pairs. Required when restricted = TRUE, otherwise optional

alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for p10 - p01 > -margin, 'less' for p10 - p01 < margin, 'two.sided' for p10 != p01, or 'equivalence' for -lower < p10 - p01 < upper
n.grid	Number of grid points used to search over the probability of a discordant pair in the test 'McNemar-uncond'. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the test 'McNemar-uncond'. The default of 0 disables the procedure
ss.method	Re-estimation rule. 'exact' (default) and 'standard' re-estimate the number of pairs of PairedSampleSize with the method of that name. 'conditional.power' takes the smallest number of pairs at which the conditional power given the interim number of discordant pairs attains tar.power, the rule of Schepers et al. (2026); it requires N.max. See Details
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
search	Search of the exact re-estimation, see PairedSampleSize; it does not go beyond N.max
search.limit	Limit of the search 'stable', see PairedSampleSize
N.min	Smallest final number of pairs, or NULL
N.max	Largest final number of pairs, or NULL
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

The re-estimation is that of PairedPowerBSSR: the blinded estimate D1 / n1 of the probability of a discordant pair, raised to |Delta.A| when it is smaller, and the assumed difference Delta.A give the planning probabilities, from which the number of pairs follows by the rule ss.method. The conditional power is sum over D2 of b(D2; N.final - n1, psi.plan) g(D1 + D2), the probability of rejection given D1 under the planning probabilities, formula (4) of Schepers et al. (2026) for the exact conditional McNemar test.

Value

An object of class bbssr_pbssr, a data frame with one row containing the interim data n1 and D1, the blinded estimate psi.hat of the probability of a discordant pair, the value psi.plan used in the re-estimation, the planning probabilities p10.plan and p01.plan, the re-estimated number of pairs

N.re before the bounds are applied (at most N.max with ss.method = 'exact', whose search does not go beyond it, and within the bounds with ss.method = 'conditional.power'), the number n.stage2 of pairs still to be enrolled, the final number of pairs N.final, the power Power of the final analysis with N.final pairs under the planning probabilities, and its conditional power CP given D1, computed with ss.Test at the level ss.alpha. The attributes hold the design settings.

References

Schepers M, Brannath W, Hoffmann E, Stingl J, Schmidtmann I (2026). Blinded sample size review for McNemar's test based on primary and surrogate endpoints. arXiv:2608.17784.

Examples

```
# 25 interim pairs of which 5 are discordant, assumed difference 0.15
PairedBSSR(n1 = 25, D1 = 5, Delta.A = 0.15, alpha = 0.05, tar.power = 0.8,
           Test = 'McNemar-exact', alternative = 'two.sided',
           ss.method = 'conditional.power', N.max = 150)
```

PairedPower *Power for Paired Binary Data*

Description

Calculates the exact power of a test comparing two treatments on a binary endpoint in a matched-pairs design, for given probabilities of the two kinds of discordant pairs.

Usage

```
PairedPower(
  p10,
  p01,
  N,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

p10	Probability that only treatment 1 responds in a pair. A vector is allowed
p01	Probability that only treatment 2 responds in a pair, of the same length as p10. A single value of either is used with every value of the other

N	Number of pairs
alpha	Level of significance for the alternative specified by alternative
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'
alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for p10 - p01 > -margin, 'less' for p10 - p01 < margin, 'two.sided' for p10 != p01, or 'equivalence' for -lower < p10 - p01 < upper
n.grid	Number of grid points used to search over the probability of a discordant pair in the test 'McNemar-uncond'. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the test 'McNemar-uncond'. The default of 0 disables the procedure
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

The numbers of the two kinds of discordant pairs and of concordant pairs follow the multinomial distribution with N trials and probabilities p10, p01 and 1 - p10 - p01, and the power is the probability of the rejection region of PairedRR under this distribution, computed by complete enumeration. The probabilities of the concordant pairs with both and with neither treatment responding do not enter, since no test depends on how the concordant pairs split.

Value

An object of class bbssr_ppower, a data frame with one row for each element of p10, containing:

p10, p01 Probabilities of the two kinds of discordant pairs

N Number of pairs

alpha Level of significance

Test Name of the statistical test

alternative Direction of the alternative hypothesis

Power Exact power

The attribute margin holds the margin.

Examples

```
# Average exact power of the asymptotic McNemar test with 25 pairs over the alternatives
# of Berger and Sidik (2003, Section 5.1)
alt <- do.call(rbind, lapply(seq(0.025, 0.475, by = 0.05), function(q) {
  data.frame(p01 = q, p10 = seq(q + 0.05, 1 - q, by = 0.05))
}))
pw <- PairedPower(alt$p10, alt$p01, N = 25, alpha = 0.05, Test = 'McNemar')
mean(pw$Power)
```

PairedPowerBSSR *Power of Blinded Sample Size Re-Estimation for Paired Binary Data*

Description

Calculates the exact power and the distribution of the final number of pairs of a matched-pairs design with blinded sample size re-estimation, together with the power of the fixed design with the planned number of pairs. The number of pairs is re-estimated at an interim analysis from the number of discordant pairs, without the direction in which they are discordant, and the final analysis uses all pairs.

Usage

```
PairedPowerBSSR(
  psi,
  Delta.T,
  Delta.A,
  N,
  omega = NULL,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  ss.method = c("exact", "standard", "conditional.power"),
  ss.Test = Test,
  ss.alpha = alpha,
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

psi	True probability p10 + p01 of a discordant pair. A vector is allowed
Delta.T	True difference p10 - p01, a single value
Delta.A	Assumed difference p10 - p01 used in the re-estimation
N	Planned number of pairs, which is the number of pairs of the fixed design
omega	Fraction of N at the interim analysis, rounded up. Supply either omega or n.interim
alpha	Level of significance for the alternative specified by alternative
tar.power	Target power of the re-estimation
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'
restricted	Logical. If TRUE, the final number of pairs is at least N. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for p10 - p01 > -margin, 'less' for p10 - p01 < margin, 'two.sided' for p10 != p01, or 'equivalence' for -lower < p10 - p01 < upper
n.grid	Number of grid points used to search over the probability of a discordant pair in the test 'McNemar-uncond'. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the test 'McNemar-uncond'. The default of 0 disables the procedure
ss.method	Re-estimation rule. 'exact' (default) and 'standard' re-estimate the number of pairs of PairedSampleSize with the method of that name. 'conditional.power' takes the smallest number of pairs at which the conditional power given the interim number of discordant pairs attains tar.power, the rule of Schepers et al. (2026); it requires N.max. See Details
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
search	Search of the exact re-estimation, see PairedSampleSize; it does not go beyond N.max
search.limit	Limit of the search 'stable', see PairedSampleSize
N.min	Smallest final number of pairs, or NULL
N.max	Largest final number of pairs, or NULL
n.interim	Number of pairs at the interim analysis. Supply either omega or n.interim
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

Among the n1 pairs of the interim analysis, the number D1 of discordant pairs is binomial with n1 trials and probability psi, and the blinded estimate of psi is D1 / n1. The assumed difference Delta.A is kept, and the estimate is raised to |Delta.A| when it is smaller, since psi >= |p10 - p01|. The planning probabilities are then p10 = (psi.plan + Delta.A) / 2 and p01 = (psi.plan - Delta.A) / 2, as in Xu and Yu (2013) and Schepers et al. (2026).

The final number of pairs depends on the interim data only through D1. Given the numbers of discordant pairs of the two stages, the number of pairs in which only treatment 1 responds is binomial with the total number of discordant pairs and probability p10 / psi, whatever the re-estimation did. The power is therefore sum over D1 and D2 of b(D1; n1, psi) b(D2; N(D1) - n1, psi) g(D1 + D2), where g(D) is the probability that the final analysis with N(D1) pairs rejects given D discordant pairs. It is computed exactly by enumeration.

With ss.method = 'conditional.power' the final number of pairs is the smallest number N between the bounds at which sum over D2 of b(D2; N - n1, psi.plan) g_N(D1 + D2), computed with the planning probabilities and the test ss.Test at the level ss.alpha, attains tar.power. With the exact conditional McNemar test this is the conditional power (4) of Schepers et al. (2026), which they evaluate with p* = 0.5 + Delta / (2 psi) at the blinded estimate of psi. When no number up to N.max attains the target, N.max is used, a choice the article leaves open.

The exact conditional McNemar test keeps its level whatever blinded rule is used, since the number of pairs in which only treatment 1 responds stays binomial given the total number of discordant pairs (Schepers et al., 2026). The other tests can exceed their nominal level after re-estimation, see PairedTypeIErrorBSSR.

Value

An object of class bbssr_ppowerbssr, a data frame with one row for each element of psi that gives probabilities p10 and p01 in [0, 1], containing:

psi, p10, p01 True probabilities of a discordant pair and of its two kinds

power.BSSR Power of the design with re-estimation

power.TRAD Power of the fixed design with N pairs

E.N Expected final number of pairs

The attribute reestimation holds the final number of pairs for every interim number of discordant pairs, N.dist the distribution of the final number of pairs in each scenario, and the other attributes the design settings.

References

Schepers M, Brannath W, Hoffmann E, Stingl J, Schmidtmann I (2026). Blinded sample size review for McNemar's test based on primary and surrogate endpoints. arXiv:2608.17784.

Xu J, Yu M (2013). Sample size determination and re-estimation for matched pair designs with multiple binary endpoints. *Biometrical Journal*, 55(3), 430-443.

Examples

```
res <- PairedPowerBSSR(psi = c(0.15, 0.2, 0.3), Delta.T = 0.15, Delta.A = 0.15, N = 72,
                       n.interim = 30, alpha = 0.05, tar.power = 0.8,
                       Test = 'McNemar-exact', alternative = 'two.sided',
                       ss.method = 'standard')
print(res)


# Conditional power rule of Schepers et al. (2026) with at most 150 pairs
res <- PairedPowerBSSR(psi = c(0.15, 0.2, 0.3), Delta.T = 0.15, Delta.A = 0.15, N = 72,
                       n.interim = 30, alpha = 0.05, tar.power = 0.8,
                       Test = 'McNemar-exact', alternative = 'two.sided',
                       ss.method = 'conditional.power', N.max = 150)
print(res)
```

PairedRR *Rejection Region for Paired Binary Data*

Description

Provides the rejection region of a test comparing two treatments on a binary endpoint in a matched-pairs design, such as a cross-over trial or a comparison of two diagnostic procedures on the same patients. Six tests are supported. Each can be applied with a one-sided or a two-sided alternative, and the score test also tests non-inferiority and equivalence with a margin on the scale of the difference of the response probabilities.

Usage

```
PairedRR(
  N,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

N	Number of pairs
alpha	Level of significance for the alternative specified by alternative
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'

`alternative` Direction of the alternative hypothesis. Options: `'greater'` (default) for `p10 - p01 > -margin`, `'less'` for `p10 - p01 < margin`, `'two.sided'` for `p10 != p01`, or `'equivalence'` for `-lower < p10 - p01 < upper`

`n.grid` Number of grid points used to search over the probability of a discordant pair in the test `'McNemar-uncond'`. Default is 100

`bb.gamma` Confidence level parameter of the Berger-Boos procedure for the test `'McNemar-uncond'`. The default of 0 disables the procedure

`margin` Margin on the scale of the difference `p10 - p01`. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis `p10 - p01 <= -margin` against `p10 - p01 > -margin` when `alternative` is `'greater'`, and `p10 - p01 >= margin` against `p10 - p01 < margin` when it is `'less'`. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires `Test = 'Paired-Score'` and a one-sided alternative. For `alternative = 'equivalence'` the margin is the pair `c(lower, upper)`, both in (0, 1), or a single value used for both, and the null hypothesis `p10 - p01 <= -lower` or `p10 - p01 >= upper` is tested

`ref.pvalue` Logical. If `TRUE`, the maximization over the probability of a discordant pair in the test `'McNemar-uncond'` is refined between the grid points, see Details. Default is `FALSE`

Details

A pair contributes one of four outcomes: both treatments respond, only treatment 1 responds (probability `p10`), only treatment 2 responds (probability `p01`), or neither responds. The difference of the response probabilities of the two treatments is `p10 - p01`, and only the numbers `a` and `b` of the two kinds of discordant pairs enter the tests. The function supports the following tests:

- The asymptotic McNemar test (McNemar), which refers `Z = (a - b) / sqrt(a + b)` to the standard normal distribution without a continuity correction, with `Z = 0` when `a + b = 0`
- The asymptotic McNemar test with a continuity correction (McNemar-cc), formula (2) of Fagerland et al. (2013), which refers `(|a - b| - 1) / sqrt(a + b)` to the standard normal distribution, and `(a - b - 1) / sqrt(a + b)` for the alternative `'greater'`, with the p-value 1 when `a + b = 0`
- The exact conditional McNemar test (McNemar-exact), which refers `a` to the binomial distribution with `a + b` trials and probability 1/2
- The mid-p version of the exact conditional test (McNemar-midP), formulae (5) and (6) of Fagerland et al. (2013)
- The exact unconditional test of Suissa and Shuster (1991) (McNemar-uncond), which orders the outcomes by `Z` and maximizes the null tail probability over the probability `psi = p10 + p01` of a discordant pair, the test Z of Berger and Sidik (2003)
- The score test of Nam (1997) and Tango (1998) for the null hypothesis `p10 - p01 = -margin` (Paired-Score), which refers `(a - b + N margin) / sqrt(N (2 q01 - margin (margin + 1)))` to the standard normal distribution, where `q01` is the maximum likelihood estimate of `p01` under the null hypothesis, formulae (24) to (26) of Tango (1998). With `margin = 0` it is the asymptotic McNemar test

The conditional distribution of a given a + b is symmetric under the null hypothesis, so the usual conventions for the two-sided p-value of a conditional test agree. The two-sided p-value of the exact conditional test is twice the smaller tail probability, at most 1, and the two-sided mid-p value follows formulae (5) and (6) of Fagerland et al. (2013). The two-sided versions of 'McNemar', 'McNemar-uncond' and 'Paired-Score' order the outcomes by the absolute value of the statistic. The two-sided p-value of 'McNemar-cc' is twice the smaller of its two one-sided p-values, at most 1, which is the p-value of stats::mcnemar.test with correct = TRUE. Its one-sided p-value is the normal approximation with a continuity correction to the tail probability of the exact conditional test, and Fagerland et al. (2013) set the p-value to 1 when a + b = 0, where the statistic is not defined.

The test 'McNemar-uncond' maximizes the tail probability of each outcome, a polynomial of degree N in psi, over a grid of n.grid points. The maximum on the grid can fall below the maximum over the unit interval. With ref.pvalue = TRUE the grid is extended by points equally spaced on the arcsine square-root scale and every local maximum on the grid is refined by a golden-section search between its two neighbouring grid points, as in BinaryRR. With a positive bb.gamma the maximization is restricted to the Clopper-Pearson interval of level 1 - bb.gamma for psi computed from a + b, and bb.gamma is added, the confidence interval p-value of Berger and Sidik (2003).

The alternative 'less' is handled by exchanging a and b. With alternative = 'equivalence' the null hypothesis is tested by two one-sided score tests, against 'greater' with the margin lower and against 'less' with the margin upper, each at the level alpha, and the p-value of an outcome is the larger of the two one-sided p-values.

The p-values are kept for the rest of the session and reused by later calls with the same number of pairs and test, see bbssr-package. They form a matrix of (N + 1)^2 values, which takes about 2 GB of memory for 16,000 pairs, so the memory and the time of the exact calculations grow with the square of N.

Value

An object of class bbssr_prr, a logical matrix of dimension (N + 1) by (N + 1) whose entry [a + 1, b + 1] is TRUE when the null hypothesis is rejected with a pairs in which only treatment 1 responds and b pairs in which only treatment 2 responds. Entries with a + b > N are impossible outcomes and are FALSE. The p-values are stored in the attribute p.value, with NA at the impossible outcomes, and the design settings as further attributes

References

Berger RL, Sidik K (2003). Exact unconditional tests for a 2 x 2 matched-pairs design. *Statistical Methods in Medical Research*, 12(2), 91-108.

Fagerland MW, Lydersen S, Laake P (2013). The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. *BMC Medical Research Methodology*, 13, 91.

Nam J (1997). Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics*, 53(4), 1422-1430.

Suissa S, Shuster JJ (1991). The 2 x 2 matched-pairs trial: exact unconditional design and analysis. *Biometrics*, 47(2), 361-372.

Tango T (1998). Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine*, 17(8), 891-908.

Examples

```
RR <- PairedRR(N = 20, alpha = 0.025, Test = 'McNemar-exact')
print(RR)

# Non-inferiority with a margin of 0.1 by the score test
RR <- PairedRR(N = 44, alpha = 0.05, Test = 'Paired-Score', margin = 0.1)
sum(RR)


# Exact unconditional test with the Berger-Boos procedure
RR <- PairedRR(N = 30, alpha = 0.05, Test = 'McNemar-uncond', bb.gamma = 0.0005)
plot(RR)
```

PairedSampleSize *Sample Size for Paired Binary Data*

Description

Calculates the number of pairs required for a test comparing two treatments on a binary endpoint in a matched-pairs design to attain a target power. The number can be obtained from the exact power of the selected test or from the normal approximation.

Usage

```
PairedSampleSize(
  p10,
  p01,
  alpha,
  tar.power,
  Test,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  method = c("exact", "standard"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

p10	Probability that only treatment 1 responds in a pair
p01	Probability that only treatment 2 responds in a pair
alpha	Level of significance for the alternative specified by alternative

tar.power	Target power
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'
alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for p10 - p01 > -margin, 'less' for p10 - p01 < margin, 'two.sided' for p10 != p01, or 'equivalence' for -lower < p10 - p01 < upper
n.grid	Number of grid points used to search over the probability of a discordant pair in the test 'McNemar-uncond'. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the test 'McNemar-uncond'. The default of 0 disables the procedure
method	How the number of pairs is obtained. 'exact' (default) steps from the normal approximation to the number of pairs at which the exact power of Test attains the target, see Details. 'standard' rounds up the normal approximation, formula (9) of Nam (1997), which is formula (3) of Connor (1987) when the margin is 0
search	How the search of method = 'exact' chooses the number of pairs. 'crossing' (default) steps from the normal approximation to a number that attains the target power while the number one smaller does not. 'smallest' returns the smallest number that attains the target power. 'stable' returns the smallest number from which every number up to a limit attains it. Ignored by method = 'standard'
search.limit	Two numbers a and b giving the limit max(ceiling(a n0), ceiling(n0 + b)) of the number of pairs examined by search = 'stable', where n0 is the normal approximation rounded up. a must be at least 1 and b non-negative. Default is c(2, 50)
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

The normal approximation is formula (9) of Nam (1997), n = (z.a sqrt(v0) + z.b sqrt(v1))^2 / (d1 - d0)^2, where d1 = p10 - p01 is the true difference, d0 = -margin the difference under the null hypothesis, v1 = p10 + p01 - d1^2 the variance under the alternative, and v0 = 2 p01.bar + d0 - d0^2 the variance under the null hypothesis evaluated at the large sample value p01.bar of the restricted maximum likelihood estimate of p01, formula (5) of that article. z.a is the upper alpha quantile of the standard normal distribution, the upper alpha / 2 quantile for a two-sided alternative, and z.b the tar.power quantile. With margin = 0 the formula is formula (3) of Connor

(1987). For `alternative = 'less'` the probabilities `p10` and `p01` are exchanged. For `alternative = 'equivalence'` the approximate power of each one-sided test is obtained in the same way for its own margin, at the level `alpha`, and the normal approximation is the smallest number of pairs at which the sum of the two approximate powers minus 1 attains the target power. The sum minus 1 is a lower bound of the approximate probability that both tests reject.

The normal approximation does not depend on the test, and for `'McNemar-cc'` it leaves out the continuity correction. It is the starting point of the exact search for every test. The exact power is not monotone in the number of pairs, so the three searches can return different numbers, see `BinarySampleSize`.

Value

An object of class `bbssr_psamplesize`, a data frame with one row containing:

p10, p01 Probabilities of the two kinds of discordant pairs

alpha Level of significance

tar.power Target power

Test Name of the statistical test

alternative Direction of the alternative hypothesis

Power Exact power of `Test` with the required number of pairs

N Required number of pairs

The attributes hold the remaining settings, among them the unrounded normal approximation `n.raw` and the limit `search.limit` of the search `'stable'` (otherwise `NA`).

References

Connor RJ (1987). Sample size for testing differences in proportions for the paired-sample design. *Biometrics*, 43(1), 207-211.

Nam J (1997). Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics*, 53(4), 1422-1430.

Examples

```
# Example of Connor (1987, p. 210): theta1 = 0.8, theta2 = 0.7 and psi = 0.3
PairedSampleSize(p10 = 0.2, p01 = 0.1, alpha = 0.05, tar.power = 0.9,
                 Test = 'McNemar', method = 'standard')


# Exact number of pairs of the exact conditional test
PairedSampleSize(p10 = 0.2, p01 = 0.1, alpha = 0.05, tar.power = 0.9,
                 Test = 'McNemar-exact')
```

PairedTypeIErrorBSSR *Type I Error Rate of Blinded Sample Size Re-Estimation for Paired Binary Data*

Description

Calculates the exact type I error rate of a matched-pairs design with blinded sample size re-estimation, and of the fixed design with the planned number of pairs, over the probability of a discordant pair on the boundary of the null hypothesis, together with the largest rate.

Usage

```
PairedTypeIErrorBSSR(
  Delta.A,
  N,
  omega = NULL,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided", "equivalence"),
  n.grid = 100,
  bb.gamma = 0,
  ss.method = c("exact", "standard", "conditional.power"),
  ss.Test = Test,
  ss.alpha = alpha,
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  psi = seq(0, 1, by = 0.01),
  maximize = c("certified", "refined", "grid"),
  margin = 0,
  ref.pvalue = FALSE
)
```

Arguments

Delta.A	Assumed difference p10 - p01 used in the re-estimation
N	Planned number of pairs, which is the number of pairs of the fixed design
omega	Fraction of N at the interim analysis, rounded up. Supply either omega or n.interim
alpha	Level of significance for the alternative specified by alternative
tar.power	Target power of the re-estimation
Test	Type of statistical test. Options: 'McNemar', 'McNemar-cc', 'McNemar-exact', 'McNemar-midP', 'McNemar-uncond' or 'Paired-Score'

restricted	Logical. If TRUE, the final number of pairs is at least N. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default) for p10 - p01 > -margin, 'less' for p10 - p01 < margin, 'two.sided' for p10 != p01, or 'equivalence' for -lower < p10 - p01 < upper
n.grid	Number of grid points used to search over the probability of a discordant pair in the test 'McNemar-uncond'. Default is 100
bb.gamma	Confidence level parameter of the Berger-Boos procedure for the test 'McNemar-uncond'. The default of 0 disables the procedure
ss.method	Re-estimation rule. 'exact' (default) and 'standard' re-estimate the number of pairs of PairedSampleSize with the method of that name. 'conditional.power' takes the smallest number of pairs at which the conditional power given the interim number of discordant pairs attains tar.power, the rule of Schepers et al. (2026); it requires N.max. See Details
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
search	Search of the exact re-estimation, see PairedSampleSize; it does not go beyond N.max
search.limit	Limit of the search 'stable', see PairedSampleSize
N.min	Smallest final number of pairs, or NULL
N.max	Largest final number of pairs, or NULL
n.interim	Number of pairs at the interim analysis. Supply either omega or n.interim
psi	Values of the probability p10 + p01 of a discordant pair at which the rate is evaluated. Values that do not lie on the boundary of the null hypothesis, below |margin|, are dropped. Default is seq(0, 1, by = 0.01)
maximize	How the largest rate is found. 'certified' (default) takes the largest value on the grid and certifies the maximum over the interval spanned by the grid with an upper bound, 'refined' refines the largest values on the grid between neighbouring grid points, and 'grid' takes the largest value on the grid. See Details
margin	Margin on the scale of the difference p10 - p01. The default of 0 gives a test of superiority, and a value other than 0 tests the null hypothesis p10 - p01 <= -margin against p10 - p01 > -margin when alternative is 'greater', and p10 - p01 >= margin against p10 - p01 < margin when it is 'less'. A negative value tests for superiority by more than its absolute value. A margin other than 0 requires Test = 'Paired-Score' and a one-sided alternative. For alternative = 'equivalence' the margin is the pair c(lower, upper), both in (0, 1), or a single value used for both, and the null hypothesis p10 - p01 <= -lower or p10 - p01 >= upper is tested
ref.pvalue	Logical. If TRUE, the maximization over the probability of a discordant pair in the test 'McNemar-uncond' is refined between the grid points, see Details. Default is FALSE

Details

The type I error rate is evaluated on the boundary `p10 - p01 = d0` of the null hypothesis, with `d0 = 0` for a test of superiority, `d0 = -margin` for `alternative = 'greater'` and `d0 = margin` for `'less'`, as for two independent groups. For a fixed design, Berger and Sidik (2003, Theorem 1) show that the largest rate over the null hypothesis of superiority lies on this boundary for rejection regions that are Barnard convex. A test of equivalence has two boundaries, `-lower` and `upper`, and the largest rate is the larger of the maxima on the two.

On a boundary the rejection probability is a polynomial in `psi`: given the numbers of discordant pairs, the number of pairs in which only treatment 1 responds is binomial with probability `p10 / psi`, which is fixed on the boundary up to the linear reparametrization of `paired_boundary`, and the re-estimation depends on the interim data only through the number of discordant pairs. The probability is computed exactly. With `maximize = 'certified'` the Bernstein coefficients of the polynomial over the interval spanned by the values of `psi` on the boundary are bounded by subdivision, as in `BinaryTypeIErrorBSSR`, which gives an upper bound of the largest rate over the interval within `1e-12` of the maximum found.

The exact conditional McNemar test keeps its level whatever blinded rule is used (Schepers et al., 2026), so its largest rate does not exceed `alpha`.

Value

An object of class `bbssr_ptie`, a data frame with one row for each value of `psi` on the boundary, containing `psi`, the probabilities `p10` and `p01` on the boundary, and the rates `TIE.BSSR` of the design with re-estimation and `TIE.TRAD` of the fixed design. For a test of equivalence the rates on the two boundaries are stacked, with the column `boundary`. The attribute `max` holds the largest rates `TIE` of both designs, the values `psi` at which they occur, the upper bounds `bound` of the certified maximization (otherwise `NA`) and, for a test of equivalence, the boundary. The attribute `reestimation` holds the final number of pairs for every interim number of discordant pairs, and the other attributes the design settings.

References

Berger RL, Sidik K (2003). Exact unconditional tests for a 2 x 2 matched-pairs design. *Statistical Methods in Medical Research*, 12(2), 91-108.

Schepers M, Brannath W, Hoffmann E, Stingl J, Schmidtmann I (2026). Blinded sample size review for McNemar's test based on primary and surrogate endpoints. arXiv:2608.17784.

Examples

```
tie <- PairedTypeIErrorBSSR(Delta.A = 0.15, N = 72, n.interim = 30, alpha = 0.05,
                            tar.power = 0.8, Test = 'McNemar',
                            alternative = 'two.sided', ss.method = 'standard')
print(tie)


plot(tie)
```

plot.bbssr_crp *Plot Method for bbssr_crp Objects*

Description

Draws the conditional rejection probabilities of a design with blinded sample size re-estimation over the pooled responder counts of the two stages, or the type I error rate given the pooled number of interim responders.

Usage

```
## S3 method for class 'bbssr_crp'
plot(
  x,
  what = c("CRP", "CRP.total", "by.s"),
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  legend.title = NULL,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_crp
what	Quantity to draw. 'CRP' (default) and 'CRP.total' draw the corresponding column over the pairs (s, s2), shaded from white to the first colour, and outline the outcomes at which it exceeds the nominal level in the second colour. 'by.s' draws the type I error rate given s for each value of theta, which requires an object created with theta
main	Title. NULL keeps the default and NA removes it
sub	Subtitle. NULL keeps the default and NA removes it
xlab	Label of the horizontal axis. NULL keeps the default
ylab	Label of the vertical axis. NULL keeps the default
legend.title	Title of the legend. NULL keeps the default
colours	Two colours, for the shading of the probabilities and for the outline of the outcomes above the nominal level. They are not used for what = 'by.s'
base_size	Base font size passed to theme_bw()
...	Further arguments, ignored

Value

A ggplot object

Examples

```
crp <- BinaryCondRejectBSSR(
  Delta.A = 0.3, N1 = 12, N2 = 12, n.interim = c(6, 6), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Fisher', ss.method = 'standard',
  theta = c(0.3, 0.5)
)
plot(crp)
plot(crp, what = 'by.s')
```

---

plot.bbssr_grid *Plot Method for bbssr_grid Objects*

Description

Draws the power or the expected total sample size of the designs of BinaryGridBSSR against the true pooled response probability, with one colour per value of a chosen column and, optionally, one panel per value of another.

Usage

```
## S3 method for class 'bbssr_grid'
plot(
  x,
  what = c("power", "E.N"),
  colour.by = NULL,
  facet.by = NULL,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ref.line = NULL,
  legend.title = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_grid
what	Quantity to draw. 'power' (default) draws the power of the BSSR designs as solid lines and that of the fixed-sample designs as dashed lines. 'E.N' draws the expected total sample size of the BSSR designs

colour.by	Name of the column of x whose values are told apart by colour. The default of NULL uses the index design
facet.by	Name of a column of x whose values are drawn in separate panels, or NULL (default) for a single panel
main	Title. NULL keeps the default and NA removes it
sub	Subtitle. NULL keeps the default and NA removes it
xlab	Label of the horizontal axis. NULL keeps the default
ylab	Label of the vertical axis. NULL keeps the default
ref.line	Height of the dotted reference line for what = 'power'. NULL draws it at the target power when all designs share one, and NA removes it
legend.title	Title of the colour legend. NULL uses colour.by
base_size	Base font size passed to theme_bw()
...	Further arguments, ignored

Value

A ggplot object

Examples

```
design <- expand.grid(Test = c('Chisq', 'Fisher'), omega = c(0.3, 0.5),
                      stringsAsFactors = FALSE)
grid <- BinaryGridBSSR(design, p = c(0.3, 0.4, 0.5), Delta.A = 0.3, N1 = 30, N2 = 30,
                       r = 1, alpha = 0.025, tar.power = 0.8, ss.method = 'standard')
plot(grid, colour.by = 'Test', facet.by = 'omega')
plot(grid, what = 'E.N', colour.by = 'omega', facet.by = 'Test')
```

---

plot.bbssr_power *Plot Exact Power against the Response Probability of Group 1*

Description

Plot Exact Power against the Response Probability of Group 1

Usage

```
## S3 method for class 'bbssr_power'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ylim = c(0, 1),
```

```
  show.points = TRUE,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_power returned by BinaryPower
main	Title of the plot. The default of NULL builds the title from the object and a single NA drops it
sub	Subtitle of the plot, following the convention of main
xlab	Label of the horizontal axis, following the convention of main
ylab	Label of the vertical axis, following the convention of main
ylim	Numeric vector of length two giving the range of the vertical axis, or NULL for a range chosen from the data
show.points	Logical. If TRUE, the default, a marker is drawn at every evaluated response probability. Set it to FALSE to draw the curve alone, which is easier to read on a fine grid
colours	Colour of the curve
base_size	Base font size of the theme, in points
...	Further arguments, currently ignored

Details

The vertical range is imposed with coord_cartesian, so points outside ylim are hidden rather than removed from the data.

Value

A ggplot object

Examples

```
pw <- BinaryPower(p1 = seq(0.3, 0.8, by = 0.1), p2 = rep(0.2, 6),
                  N1 = 20, N2 = 20, alpha = 0.025, Test = 'Chisq')
plot(pw)

# Enlarge the type, rescale the vertical axis, rename the axes and drop the markers
plot(pw, base_size = 14, ylim = c(0.2, 1), sub = NA, show.points = FALSE,
     xlab = 'Response probability, experimental group', ylab = 'Exact power')
```

`plot.bbssr_powerbssr` *Plot the Power of a BSSR Design against the Fixed-Sample Design*

Description

Plot the Power of a BSSR Design against the Fixed-Sample Design

Usage

```
## S3 method for class 'bbssr_powerbssr'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ylim = NULL,
  ref.line = NULL,
  legend.title = NULL,
  legend.labels = NULL,
  show.points = TRUE,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

`x`	An object of class `bbssr_powerbssr` returned by `BinaryPowerBSSR`
`main`	Title of the plot. The default of `NULL` builds the title from the object and a single `NA` drops it
`sub`	Subtitle of the plot, following the convention of `main`
`xlab`	Label of the horizontal axis, following the convention of `main`
`ylab`	Label of the vertical axis, following the convention of `main`
`ylim`	Numeric vector of length two giving the range of the vertical axis, or `NULL` for a range chosen from the data
`ref.line`	Numeric vector of heights at which a dashed horizontal line is drawn. The default of `NULL` places one line at the target power and a single `NA` draws no line
`legend.title`	Title of the legend. The default of `NULL` leaves the legend untitled
`legend.labels`	Character vector of length two replacing the entries of the legend, in the order BSSR then fixed sample
`show.points`	Logical. If `TRUE`, the default, a marker is drawn at every evaluated pooled probability. Set it to `FALSE` to draw the curves alone, which is easier to read on a fine grid

colours	Character vector of length two giving the colour of the two curves, in the order BSSR then fixed sample
base_size	Base font size of the theme, in points
...	Further arguments, currently ignored

Details

Setting Delta.T to 0 in BinaryPowerBSSR turns the two power columns into rejection probabilities under the null hypothesis. The default reference line at the target power is then out of place, and the level of significance passed through ref.line together with a rescaled ylim gives a readable plot of the type I error rate.

The vertical range is imposed with coord_cartesian, so points outside ylim are hidden rather than removed from the data.

Value

A ggplot object

Examples

```
res <- BinaryPowerBSSR(
  p = seq(0.23, 0.47, by = 0.04),
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 39, N2 = 39, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq'
)
plot(res)

# Type I error rate, with the reference line moved to the level of significance
tie <- BinaryPowerBSSR(
  p = seq(0.23, 0.47, by = 0.04),
  Delta.A = 0.3, Delta.T = 0,
  N1 = 39, N2 = 39, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq'
)
plot(tie, ref.line = 0.025, ylim = c(0.015, 0.030), base_size = 14,
     show.points = FALSE,
     main = 'Type I error rate of the BSSR design', ylab = 'Type I error rate',
     legend.title = 'Design', legend.labels = c('BSSR', 'Fixed'))
```

---

plot.bbssr_prr *Plot a Rejection Region for Paired Binary Data*

Description

Displays the possible outcomes of a matched-pairs trial and shades the outcomes for which the null hypothesis is rejected.

Usage

```
## S3 method for class 'bbssr_prr'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  legend.title = NULL,
  legend.labels = NULL,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_prr returned by PairedRR
main	Title of the plot. The default of NULL builds the title from the object and a single NA drops it
sub	Subtitle of the plot, following the convention of main
xlab	Label of the horizontal axis, following the convention of main
ylab	Label of the vertical axis, following the convention of main
legend.title	Title of the legend. The default of NULL leaves the legend untitled
legend.labels	Character vector of length two replacing the entries of the legend, in the order retained then rejected
colours	Character vector of length two giving the fill of the tiles, in the order retained then rejected
base_size	Base font size of the theme, in points
...	Further arguments, currently ignored

Details

The vertical axis counts the pairs in which only treatment 1 responds and runs downwards, and the horizontal axis counts the pairs in which only treatment 2 responds, as in the map printed by print.bbssr_prr. Only the possible outcomes, with at most N discordant pairs, are drawn.

Value

A ggplot object

Examples

```
RR <- PairedRR(N = 20, alpha = 0.025, Test = 'McNemar-midP')
plot(RR)
```

plot.bbssr_ptie *Plot Method for bbssr_ptie Objects*

Description

Draws the type I error rate of a matched-pairs design with blinded sample size re-estimation and of the fixed design against the probability of a discordant pair on the boundary of the null hypothesis, with a reference line at the nominal level. For a test of equivalence the two boundaries are drawn in separate panels.

Usage

```
## S3 method for class 'bbssr_ptie'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ylim = NULL,
  ref.line = NULL,
  legend.title = NULL,
  legend.labels = NULL,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_ptie
main	Title. NULL keeps the default and NA removes it
sub	Subtitle. NULL keeps the default and NA removes it
xlab	Label of the horizontal axis. NULL keeps the default
ylab	Label of the vertical axis. NULL keeps the default
ylim	Range of the vertical axis, applied through coord_cartesian()
ref.line	Height of the dashed reference line. NULL draws it at the nominal level and NA removes it
legend.title	Title of the legend. NULL keeps the default
legend.labels	Labels of the two designs. NULL keeps the defaults
colours	Two colours for the BSSR and the fixed design
base_size	Base font size passed to theme_bw()
...	Further arguments, ignored

Value

A ggplot object

Examples

```
tie <- PairedTypeIErrorBSSR(Delta.A = 0.2, N = 40, n.interim = 20, alpha = 0.05,
                            tar.power = 0.8, Test = 'McNemar',
                            alternative = 'two.sided', ss.method = 'standard',
                            psi = seq(0.1, 0.9, by = 0.05))
plot(tie)
```

---

plot.bbssr_rr *Plot a Rejection Region*

Description

Displays the outcome grid of a two-arm trial with a binary endpoint and shades the outcomes for which the null hypothesis is rejected.

Usage

```
## S3 method for class 'bbssr_rr'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  legend.title = NULL,
  legend.labels = NULL,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_rr returned by BinaryRR
main	Title of the plot. The default of NULL builds the title from the object and a single NA drops it
sub	Subtitle of the plot, following the convention of main
xlab	Label of the horizontal axis, following the convention of main
ylab	Label of the vertical axis, following the convention of main
legend.title	Title of the legend. The default of NULL leaves the legend untitled

legend.labels	Character vector of length two replacing the entries of the legend, in the order retained then rejected
colours	Character vector of length two giving the fill of the tiles, in the order retained then rejected
base_size	Base font size of the theme, in points
...	Further arguments, currently ignored

Details

Both axes count responders, so the breaks are restricted to integers. The vertical axis runs downwards, which places the outcome with no responders in the top left corner and matches the layout of the map printed by print.bbssr_rr.

Value

A ggplot object

Examples

```
RR <- BinaryRR(N1 = 10, N2 = 10, alpha = 0.025, Test = 'Chisq')
plot(RR)

# Enlarge the type and relabel the legend
plot(RR, base_size = 14, legend.title = 'Decision',
     legend.labels = c('do not reject', 'reject'),
     colours = c('white', 'grey30'))
```

---

plot.bbssr_samplesize *Plot the Exact Power Curve around a Sample Size Solution*

Description

Recomputes the exact power over a range of sample sizes of group 2 and marks the selected sample size and the target power.

Usage

```
## S3 method for class 'bbssr_samplesize'
plot(
  x,
  N2.range = NULL,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ylim = NULL,
```

```
  ref.line = NULL,
  ref.line.N2 = NULL,
  show.points = TRUE,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_samplesize returned by BinarySampleSize
N2.range	Optional integer vector of sample sizes of group 2 at which the power is evaluated. By default the selected sample size plus or minus five is used
main	Title of the plot. The default of NULL builds the title from the object and a single NA drops it
sub	Subtitle of the plot, following the convention of main
xlab	Label of the horizontal axis, following the convention of main
ylab	Label of the vertical axis, following the convention of main
ylim	Numeric vector of length two giving the range of the vertical axis, or NULL for a range chosen from the data
ref.line	Numeric vector of heights at which a dashed horizontal line is drawn. The default of NULL places one line at the target power and a single NA draws no line
ref.line.N2	Numeric vector of positions at which a dotted vertical line is drawn. The default of NULL places one line at the selected sample size and a single NA draws no line
show.points	Logical. If TRUE, the default, a marker is drawn at every evaluated sample size. Set it to FALSE to draw the curve alone, which is easier to read over a wide range
colours	Colour of the curve
base_size	Base font size of the theme, in points
...	Further arguments, currently ignored

Details

The power is recomputed at every point of N2.range, so a wide range combined with one of the unconditional tests can take a long time to evaluate.

The vertical range is imposed with coord_cartesian, so points outside ylim are hidden rather than removed from the data.

Value

A ggplot object

Examples

```
ss <- BinarySampleSize(p1 = 0.4, p2 = 0.2, r = 1, alpha = 0.025,
                       tar.power = 0.8, Test = 'Chisq')
plot(ss)

# Enlarge the type, rename the axes and draw neither reference line
plot(ss, base_size = 14, ref.line = NA, ref.line.N2 = NA,
     xlab = 'Sample size of the control group', ylab = 'Exact power')
```

`plot.bbssr_sgrid` *Plot Method for bbssr_sgrid Objects*

Description

Draws the power or the expected total sample size of the stratified designs of `StratGridBSSR` against a chosen column, with one colour per value of another column and, optionally, one panel per value of a third, as `plot.bbssr_grid` does for two groups.

Usage

```
## S3 method for class 'bbssr_sgrid'
plot(x, x.by = NULL, xlab = NULL, ...)
```

Arguments

`x`	An object of class `bbssr_sgrid`
`x.by`	Name of the numeric column of `x` on the horizontal axis. The default of `NULL` takes the first of the true pooled response probabilities `p.S1`, `p.S2`, `...` that varies between the scenarios, or the index `scenario` when none varies
`xlab`	Label of the horizontal axis. `NULL` makes it from `x.by`
`...`	Further arguments of `plot.bbssr_grid`: `what`, `colour.by`, `facet.by`, `main`, `sub`, `ylab`, `ref.line`, `legend.title` and `base_size`

Details

The lines join the scenarios of a design in the order of the horizontal axis, so the scenarios should differ in a single probability, or `x.by` should name a column in which they all differ; a warning is given when two scenarios of a design share a value on the axis.

Value

A `ggplot` object

Examples

```
design <- data.frame(Test = c('CMH', 'Strat-Fisher'))
grid <- StratGridBSSR(design, p = rbind(c(0.3, 0.5), c(0.4, 0.6)), Delta.A = 0.3,
                      N1 = c(10, 10), N2 = c(10, 10), omega = 0.5, r = 1,
                      alpha = 0.05, tar.power = 0.8, ss.method = 'standard')
plot(grid, colour.by = 'Test')
plot(grid, what = 'E.N', x.by = 'p.S2', colour.by = 'Test')
```

---

plot.bbssr_spowerbssr *Plot the Power of a Stratified BSSR Design against the Fixed Design*

Description

Draws the power of the design with blinded sample size re-estimation of StratPowerBSSR and that of the fixed design against a chosen column of the scenarios, as plot.bbssr_powerbssr does for two groups.

Usage

```
## S3 method for class 'bbssr_spowerbssr'
plot(x, x.by = NULL, xlab = NULL, ...)
```

Arguments

x	An object of class bbssr_spowerbssr returned by StratPowerBSSR
x.by	Name of the numeric column of x on the horizontal axis, or 'scenario' for the rows of p that the scenarios come from. The default of NULL takes the first of the true pooled response probabilities p.S1, p.S2, ... that varies between the scenarios, or the scenario when none varies
xlab	Label of the horizontal axis. NULL makes it from x.by and a single NA drops it
...	Further arguments of plot.bbssr_powerbssr: main, sub, ylab, ylim, ref.line, legend.title, legend.labels, show.points, colours and base_size

Details

The lines join the scenarios in the order of the horizontal axis, so the scenarios should differ in a single probability, or x.by should name a column in which they all differ; a warning is given when two scenarios share a value on the axis.

The Monte Carlo standard errors of a simulated or averaged power (three or more strata, or a random allocation) are not drawn. With Delta.T = 0 in StratPowerBSSR the curves are rejection probabilities under the null hypothesis, and ref.line can be set to the level of significance.

Value

A ggplot object

Examples

```
res <- StratPowerBSSR(p = cbind(0.4, seq(0.3, 0.7, by = 0.1)), Delta.A = 0.3,
                      Delta.T = 0.3, N1 = c(10, 10), N2 = c(10, 10), omega = 0.5,
                      r = 1, alpha = 0.05, tar.power = 0.8, Test = 'CMH',
                      ss.method = 'standard')
plot(res)
```

---

plot.bbssr_stie *Plot Method for bbssr_stie Objects*

Description

Plots the type I error rate of a two-stratum design with blinded sample size re-estimation and of the fixed design over the grid of common response probabilities of the two strata, as two heat maps.

Usage

```
## S3 method for class 'bbssr_stie'
plot(
  x,
  main = NULL,
  xlab = NULL,
  ylab = NULL,
  low = "white",
  high = "firebrick",
  base_size = 11,
  ...
)
```

Arguments

x	An object of class bbssr_stie with two strata, computed on a grid
main	Title, or NULL for the default
xlab	Label of the horizontal axis, or NULL for the default
ylab	Label of the vertical axis, or NULL for the default
low	Colour of the smallest rate. Default is 'white'
high	Colour of the largest rate. Default is 'firebrick'
base_size	Base font size. Default is 11
...	Further arguments, ignored

Value

A ggplot object

Examples

```
tie <- StratTypeIErrorBSSR(Delta.A = 0.2, N1 = c(20, 20), N2 = c(20, 20), omega = 0.5,
                           r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                           ss.method = 'standard', theta = seq(0.1, 0.9, by = 0.1))
plot(tie)
```

`plot.bbssr_tie` *Plot Method for bbssr_tie Objects*

Description

Draws the type I error rate of a design with blinded sample size re-estimation and of the corresponding fixed-sample design against the common response probability, with a reference line at the nominal level. For a test of equivalence the two boundaries of the null hypothesis are drawn in separate panels.

Usage

```
## S3 method for class 'bbssr_tie'
plot(
  x,
  main = NULL,
  sub = NULL,
  xlab = NULL,
  ylab = NULL,
  ylim = NULL,
  ref.line = NULL,
  legend.title = NULL,
  legend.labels = NULL,
  colours = NULL,
  base_size = 11,
  ...
)
```

Arguments

`x`	An object of class `bbssr_tie`
`main`	Title. `NULL` keeps the default and `NA` removes it
`sub`	Subtitle. `NULL` keeps the default and `NA` removes it
`xlab`	Label of the horizontal axis. `NULL` keeps the default
`ylab`	Label of the vertical axis. `NULL` keeps the default
`ylim`	Range of the vertical axis, applied through `coord_cartesian()`
`ref.line`	Height of the dashed reference line. `NULL` draws it at the nominal level and `NA` removes it

legend.title	Title of the legend. NULL keeps the default
legend.labels	Labels of the two designs. NULL keeps the defaults
colours	Two colours for the BSSR and the fixed-sample design
base_size	Base font size passed to theme_bw()
...	Further arguments, ignored

Value

A ggplot object

Examples

```
tie <- BinaryTypeIErrorBSSR(
  Delta.A = 0.3, N1 = 20, N2 = 20, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = seq(0.05, 0.95, by = 0.05)
)
plot(tie)
```

---

print.bbssr_alphaadj *Print Method for bbssr_alphaadj Objects*

Description

Prints the adjusted nominal significance levels of a design with blinded sample size re-estimation and of the corresponding fixed-sample design.

Usage

```
## S3 method for class 'bbssr_alphaadj'
print(x, digits = 6, ...)
```

Arguments

x	An object of class bbssr_alphaadj
digits	Number of significant digits. Default is 6
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

The adjusted levels are rounded down to digits significant digits, so that a printed level does not reject more outcomes than the level it stands for.

Value

x, invisibly

Examples

```
adj <- BinaryAlphaAdjBSSR(
  Delta.A = 0.3, N1 = 39, N2 = 39, n.interim = c(20, 20), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = seq(0.01, 0.99, by = 0.01)
)
print(adj)
```

---

print.bbssr_bssr *Print a Sample Size Re-estimation from Interim Data*

Description

Print a Sample Size Re-estimation from Interim Data

Usage

```
## S3 method for class 'bbssr_bssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_bssr returned by BinaryBSSR
digits	Number of significant digits used for the proportions and the power
...	Further arguments, currently ignored

Value

The object x, returned invisibly

Examples

```
res <- BinaryBSSR(n1 = 20, n2 = 20, S = 11, Delta.A = 0.3, r = 1,
                  alpha = 0.025, tar.power = 0.8, Test = 'Chisq')
print(res)
```

print.bbssr_crp *Print Method for bbssr_crp Objects*

Description

Prints the design settings, the number of outcomes at which the conditional rejection probabilities exceed the nominal level, their largest values and, when the type I error rate has been decomposed, its largest value over the supplied response probabilities.

Usage

```
## S3 method for class 'bbssr_crp'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_crp
digits	Number of significant digits of the probabilities. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
crp <- BinaryCondRejectBSSR(
  Delta.A = 0.3, N1 = 12, N2 = 12, n.interim = c(6, 6), r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Fisher', ss.method = 'standard',
  theta = c(0.3, 0.5)
)
print(crp)
```

print.bbssr_pbssr *Print Method for bbssr_pbssr Objects*

Description

Prints the result of a blinded sample size re-estimation from paired interim data.

Usage

```
## S3 method for class 'bbssr_pbssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_pbssr returned by PairedBSSR
digits	Number of decimal places. Default is 4
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
res <- PairedBSSR(n1 = 25, D1 = 5, Delta.A = 0.15, alpha = 0.05, tar.power = 0.8,
                  Test = 'McNemar', alternative = 'two.sided', ss.method = 'standard')
print(res)
```

---

print.bbssr_power *Print Exact Power Results*

Description

Print Exact Power Results

Usage

```
## S3 method for class 'bbssr_power'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_power returned by BinaryPower
digits	Number of significant digits used for the power
...	Further arguments, currently ignored

Value

The object x, returned invisibly

Examples

```
pw <- BinaryPower(p1 = 0.5, p2 = 0.2, N1 = 5, N2 = 5, alpha = 0.025, Test = 'Chisq')
print(pw)
```

print.bbssr_powerbssr *Print the Operating Characteristics of a BSSR Design*

Description

Print the Operating Characteristics of a BSSR Design

Usage

```
## S3 method for class 'bbssr_powerbssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_powerbssr returned by BinaryPowerBSSR
digits	Number of significant digits used for the power
...	Further arguments, currently ignored

Value

The object x, returned invisibly

Examples

```
res <- BinaryPowerBSSR(
  p = 0.45,
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 5, N2 = 5, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq'
)
print(res)
```

print.bbssr_ppower *Print Method for bbssr_ppower Objects*

Description

Prints the exact power of a test for paired binary data.

Usage

```
## S3 method for class 'bbssr_ppower'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_ppower returned by PairedPower
digits	Number of decimal places of the power. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
pw <- PairedPower(p10 = c(0.3, 0.35), p01 = 0.1, N = 40, alpha = 0.025,
                  Test = 'McNemar-exact')
print(pw)
```

---

print.bbssr_ppowerbssr

Print Method for bbssr_ppowerbssr Objects

Description

Prints the power and the expected final number of pairs of a matched-pairs design with blinded sample size re-estimation.

Usage

```
## S3 method for class 'bbssr_ppowerbssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_ppowerbssr returned by PairedPowerBSSR
digits	Number of decimal places. Default is 4
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
res <- PairedPowerBSSR(psi = c(0.2, 0.3), Delta.T = 0.15, Delta.A = 0.15, N = 72,
                       n.interim = 30, alpha = 0.05, tar.power = 0.8,
                       Test = 'McNemar', alternative = 'two.sided',
                       ss.method = 'standard')
print(res)
```

---

print.bbssr_prr *Print Method for bbssr_prr Objects*

Description

Prints the settings of a rejection region for paired binary data and, for a small number of pairs, a map of the outcomes.

Usage

```
## S3 method for class 'bbssr_prr'
print(x, show.map = NULL, ...)
```

Arguments

x	An object of class bbssr_prr returned by PairedRR
show.map	Logical. Whether the map of the outcomes is printed. The default of NULL prints it when there are at most 20 pairs
...	Further arguments, ignored

Details

The map has one row for each number a of pairs in which only treatment 1 responds and one column for each number b of pairs in which only treatment 2 responds. X marks a rejected outcome, . a retained outcome, and the impossible outcomes a + b > N are left blank.

Value

x, invisibly

Examples

```
RR <- PairedRR(N = 15, alpha = 0.05, Test = 'McNemar')
print(RR)
```

print.bbssr_psamplesize *Print Method for bbssr_psamplesize Objects*

Description

Prints the number of pairs required for a test for paired binary data.

Usage

```
## S3 method for class 'bbssr_psamplesize'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_psamplesize returned by PairedSampleSize
digits	Number of decimal places of the power. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
ss <- PairedSampleSize(p10 = 0.2, p01 = 0.1, alpha = 0.05, tar.power = 0.9,
                       Test = 'McNemar', method = 'standard')
print(ss)
```

print.bbssr_ptie *Print Method for bbssr_ptie Objects*

Description

Prints the settings and the largest type I error rates of a matched-pairs design with blinded sample size re-estimation and of the fixed design.

Usage

```
## S3 method for class 'bbssr_ptie'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_ptie returned by PairedTypeIErrorBSSR
digits	Number of significant digits of the rates. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
tie <- PairedTypeIErrorBSSR(Delta.A = 0.2, N = 40, n.interim = 20, alpha = 0.05,
                            tar.power = 0.8, Test = 'McNemar',
                            alternative = 'two.sided', ss.method = 'standard',
                            psi = seq(0.1, 0.9, by = 0.1))
print(tie)
```

---

print.bbssr_rr *Print a Rejection Region*

Description

Print a Rejection Region

Usage

```
## S3 method for class 'bbssr_rr'
print(x, show.map = NULL, ...)
```

Arguments

x	An object of class bbssr_rr returned by BinaryRR
show.map	Logical. If TRUE, the outcome grid is printed as a map in which X marks rejection of the null hypothesis. The default prints the map only when the grid has at most 400 cells
...	Further arguments, currently ignored

Value

The object x, returned invisibly

Examples

```
RR <- BinaryRR(N1 = 5, N2 = 5, alpha = 0.025, Test = 'Chisq')
print(RR)
```

print.bbssr_samplesize *Print a Sample Size Calculation*

Description

Print a Sample Size Calculation

Usage

```
## S3 method for class 'bbssr_samplesize'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_samplesize returned by BinarySampleSize
digits	Number of significant digits used for the power
...	Further arguments, currently ignored

Value

The object x, returned invisibly

Examples

```
ss <- BinarySampleSize(p1 = 0.4, p2 = 0.2, r = 1, alpha = 0.025,
                       tar.power = 0.8, Test = 'Chisq')
print(ss)
```

print.bbssr_sbssr *Print Method for bbssr_sbssr Objects*

Description

Prints the re-estimated group sizes of a stratified trial returned by StratBSSR.

Usage

```
## S3 method for class 'bbssr_sbssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_sbssr
digits	Number of decimal places. Default is 4
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
re <- StratBSSR(n1 = c(15, 15), n2 = c(15, 15), S = c(12, 18), Delta.A = 0.2, r = 1,
                alpha = 0.025, tar.power = 0.8, Test = 'CMH', N1 = c(30, 30),
                N2 = c(30, 30), ss.method = 'standard')
print(re)
```

---

print.bbssr_scrp *Print Method for bbssr_scrp Objects*

Description

Prints the design settings, the number of outcomes at which the conditional rejection probabilities of a stratified design exceed the nominal level, their largest values and, when the type I error rate has been decomposed, its largest value over the supplied response probabilities, as returned by StratCondRejectBSSR.

Usage

```
## S3 method for class 'bbssr_scrp'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_scrp
digits	Number of significant digits of the probabilities. Default is 4
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
crp <- StratCondRejectBSSR(Delta.A = 0.3, N1 = c(8, 8), N2 = c(8, 8), omega = 0.5,
                           r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                           ss.method = 'standard')
print(crp)
```

`print.bbssr_spower` *Print Method for bbssr_spower Objects*

Description

Prints the exact power of a stratified test returned by `StratPower`.

Usage

```
## S3 method for class 'bbssr_spower'
print(x, digits = 4, ...)
```

Arguments

`x`	An object of class `bbssr_spower`
`digits`	Number of decimal places. Default is 4
`...`	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

`x`, invisibly

Examples

```
pw <- StratPower(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), N1 = c(40, 40), N2 = c(40, 40),
                 alpha = 0.025, Test = 'CMH')
print(pw)
```

print.bbssr_spowerbssr *Print Method for bbssr_spowerbssr Objects*

Description

Prints the power and the expected final total sample size of a stratified design with blinded sample size re-estimation returned by StratPowerBSSR.

Usage

```
## S3 method for class 'bbssr_spowerbssr'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_spowerbssr
digits	Number of decimal places. Default is 4
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
res <- StratPowerBSSR(p = c(0.3, 0.5), Delta.A = 0.2, Delta.T = 0.2,
                      N1 = c(30, 30), N2 = c(30, 30), omega = 0.5, r = 1,
                      alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                      ss.method = 'standard')
print(res)
```

print.bbssr_srr *Print Method for bbssr_srr Objects*

Description

Prints the critical values of a stratified test returned by StratRR.

Usage

```
## S3 method for class 'bbssr_srr'
print(x, n = 20, ...)
```

Arguments

x	An object of class bbssr_srr
n	Number of rows of the table to print. Default is 20
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
rr <- StratRR(N1 = c(10, 12), N2 = c(10, 12), alpha = 0.025, Test = 'CMH')
print(rr, n = 10)
```

print.bbssr_ssamplesize *Print Method for bbssr_ssamplesize Objects*

Description

Prints the sample size of a stratified test returned by StratSampleSize.

Usage

```
## S3 method for class 'bbssr_ssamplesize'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_ssamplesize
digits	Number of decimal places of the power. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
ss <- StratSampleSize(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), frac = c(0.5, 0.5),
                      alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                      method = 'standard')
print(ss)
```

---

print.bbssr_stie *Print Method for bbssr_stie Objects*

Description

Prints the largest type I error rates of a stratified design with blinded sample size re-estimation and of the fixed design, returned by StratTypeIErrorBSSR.

Usage

```
## S3 method for class 'bbssr_stie'
print(x, digits = 6, ...)
```

Arguments

x	An object of class bbssr_stie
digits	Number of significant digits. Default is 6
...	Further arguments, ignored

Details

An object whose attributes were removed, for example by selecting some of its columns, is printed as a data frame.

Value

x, invisibly

Examples

```
tie <- StratTypeIErrorBSSR(Delta.A = 0.2, N1 = c(20, 20), N2 = c(20, 20), omega = 0.5,
                           r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                           ss.method = 'standard', theta = seq(0.1, 0.9, by = 0.2))
print(tie)
```

---

print.bbssr_tie *Print Method for bbssr_tie Objects*

Description

Prints the design settings and the largest type I error rate of a design with blinded sample size re-estimation and of the corresponding fixed-sample design.

Usage

```
## S3 method for class 'bbssr_tie'
print(x, digits = 4, ...)
```

Arguments

x	An object of class bbssr_tie
digits	Number of significant digits of the type I error rates. Default is 4
...	Further arguments, ignored

Value

x, invisibly

Examples

```
tie <- BinaryTypeIErrorBSSR(
  Delta.A = 0.3, N1 = 20, N2 = 20, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq', ss.method = 'standard',
  theta = seq(0.05, 0.95, by = 0.05)
)
print(tie)
```

`StratAlphaAdjBSSR` *Adjusted Significance Level of Blinded Sample Size Re-Estimation for a Stratified Design*

Description

Lowers the nominal significance level of a two-stratum design with blinded sample size re-estimation until the largest type I error rate over the common response probabilities of the strata does not exceed the target level `alpha`, and returns the adjusted level together with that of the fixed design with the planned group sizes.

Usage

```
StratAlphaAdjBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard"),
  ss.Test = Test,
  ss.alpha = alpha,
  ss.strata = c("stratified", "pooled"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  frac = NULL,
  theta = seq(0, 1, by = 0.05),
  maximize = c("certified", "grid"),
  adjust = c("test", "both"),
  tol = 1e-08,
  step = 1e-05,
  alloc = c("fixed", "stratum", "group", "both"),
  block = 4,
  ss.frac = c("observed", "planned"),
  nseq = 100
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, common to the strata and used to divide the blinded estimate of the pooled response probability of each stratum between the groups at the re-estimation
N1	Planned sizes of group 1 in the strata, which are those of the fixed design
N2	Planned sizes of group 2 in the strata
omega	Fraction of the planned sizes at the interim analysis: in stratum j the interim size of group 2 is ceiling(omega N2[j]) and that of group 1 is ceiling(r[j] n2[j]). Supply either omega or n.interim
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. The blinded estimate of the pooled response probability of a stratum is divided between the groups with this ratio, also when the interim sizes are not in this ratio
alpha	Target level of significance, which the type I error rate must not exceed
tar.power	Target power of the re-estimation
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
restricted	Logical. If TRUE, the final group sizes are at least the planned sizes. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference, 'RR' for the risk ratio or 'OR' for the odds ratio
ss.method	Re-estimation rule. 'exact' (default) re-estimates the sample size of StratSampleSize with method = 'exact', and 'standard' that of the normal approximation. With three or more strata only 'standard' is available
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level of significance used for the re-estimation when adjust = 'test'. Default is alpha. It is ignored when adjust = 'both'
ss.strata	Blinded estimate of the re-estimation. 'stratified' (default) estimates the pooled response probability of each stratum from its own interim patients, and 'pooled' uses the proportion of responders among all the interim patients for every stratum, as a re-estimation that ignores the strata would. With 'pooled' the final group sizes depend only on the total number of interim responders

search	Search of the exact re-estimation, see StratSampleSize; it does not go beyond the largest size that N.max allows
search.limit	Limit of the search 'stable', see StratSampleSize
N.min	Smallest final total sample size, or NULL (default)
N.max	Largest final total sample size, or NULL (default)
n.interim	Interim group sizes, as a matrix with two rows (group 1 and group 2) and one column per stratum. Supply either omega or n.interim
frac	Fractions of the strata used to divide the re-estimated total sample size between the strata, see StratSampleSize. Default is NULL, which takes the fractions of the planned design, (N1 + N2) / sum(N1 + N2)
theta	Common response probabilities of the strata under the null hypothesis. A vector gives, with two strata, the grid of all pairs of its values, and with three or more strata the points at which every stratum has the same value. A matrix with one column per stratum gives the points directly. Default is seq(0, 1, by = 0.05)
maximize	How the largest type I error rate at a level is located, as in StratTypeIErrorBSSR. With 'certified' (default) the adjusted level controls the type I error rate over the whole rectangle spanned by theta. With 'grid' the control is assessed at the points of theta
adjust	Which part of the design uses the adjusted level. 'test' (default) applies it to the final analysis only, so the re-estimation keeps using ss.alpha. 'both' applies it to the re-estimation as well
tol	Tolerance of the bisection, relative to alpha, a single value in (0, 1). Default is 1e-8
step	Decrement of the level in the search used when adjust = 'both'. Default is 1e-5
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the planned and interim group sizes. The others make the sizes random, see Details: 'stratum' (stratum sizes random, permuted blocks within the strata) and 'both' (stratum and group sizes random), with any number of strata, and 'group' (stratum sizes in the fixed fractions, group sizes random), with two strata
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4
ss.frac	Fractions of the strata used by the re-estimation with a random allocation: 'observed' (default), those of the interim patients, or 'planned', those given by frac
nseq	Number of allocation sequences drawn with a random allocation and two strata. Default is 100

Details

The searches are those of BinaryAlphaAdjBSSR. With adjust = 'test' the final group sizes do not depend on the level, the largest type I error rate is a non-decreasing step function of the level, and a bisection locates the level above which it exceeds alpha, to within tol * alpha. The reported level is the largest value with at most six significant digits that gives the same critical values. With

`adjust = 'both'` the level is lowered from `alpha` in steps of `step` until the type I error rate is controlled. Each level requires the coefficients of the type I error rate of `StratTypeIErrorBSSR`, so the search takes about as long as that many evaluations of the type I error rate.

With a random allocation (`alloc` other than `'fixed'`) the type I error rate of the design with re-estimation at each level is that of `StratTypeIErrorBSSR`, the average over the same `nseq` allocation sequences at every level, so each level takes about `nseq` times as long as with the fixed allocation. The adjusted level controls this average, whose Monte Carlo error it does not account for. The rate of the fixed design with the planned total and the same allocation is exact. Both levels are reported as the search found them, without the rounding to six significant digits, since the critical values differ between the allocation sequences and between the stratum and group sizes.

Value

An object of class `bbssr_alphaadj`, a data frame with one row for the BSSR design and one for the fixed design containing:

Design `'BSSR'` or `'TRAD'`

alpha Target level

max.TIE Largest type I error rate at the nominal level `alpha`

alpha.adj Adjusted nominal level. It is `alpha` when the nominal level already controls the type I error rate, and 0, with a warning, when no level examined by the bisection does. The search of `adjust = 'both'` for the BSSR design stops with an error instead when it reaches 0

max.TIE.adj Largest type I error rate at the adjusted level

theta.S1.adj, theta.S2.adj Common response probabilities of the strata at which `max.TIE.adj` occurs, `NA` when `alpha.adj` is 0

max.TIE.bound Upper bound of the type I error rate at the nominal level with `maximize = 'certified'`, otherwise `NA`

max.TIE.adj.bound Upper bound of the type I error rate at the adjusted level with `maximize = 'certified'`, otherwise `NA`

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

`StratTypeIErrorBSSR`

Examples

```
StratAlphaAdjBSSR(Delta.A = 0.25, N1 = c(15, 15), N2 = c(15, 15), omega = 0.5, r = 1,
                  alpha = 0.025, tar.power = 0.8, Test = 'CMH', ss.method = 'standard',
                  theta = seq(0.05, 0.95, by = 0.05))
```

StratBSSR *Blinded Sample Size Re-Estimation for a Stratified Trial*

Description

Re-estimates the final group sizes of the strata of an ongoing two-arm trial with a binary endpoint from the blinded data of an interim analysis: the numbers of patients and the pooled numbers of responders of the strata.

Usage

```
StratBSSR(
  n1,
  n2,
  S,
  Delta.A,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  N1 = NULL,
  N2 = NULL,
  frac = NULL,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard"),
  ss.Test = Test,
  ss.alpha = alpha,
  ss.strata = c("stratified", "pooled"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  alloc = c("fixed", "stratum", "group", "both"),
  block = 4,
  ss.frac = c("observed", "planned")
)
```

Arguments

n1	Numbers of patients of group 1 observed in the strata at the interim analysis. With a random allocation a group may be empty in a stratum
n2	Numbers of patients of group 2 observed in the strata at the interim analysis
S	Numbers of responders observed in the strata at the interim analysis, pooled over both groups, one element per stratum

Delta.A	Assumed treatment effect, on the scale given by effect, common to the strata and used to divide the blinded estimate of the pooled response probability of each stratum between the groups at the re-estimation
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. The blinded estimate of the pooled response probability of a stratum is divided between the groups with this ratio, also when the interim sizes are not in this ratio
alpha	Level of significance of the final analysis
tar.power	Target power of the re-estimation
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
restricted	Logical. If TRUE, the final group sizes are at least the planned sizes. Default is FALSE
N1	Planned sizes of group 1 in the strata, or NULL (default). Required when restricted is TRUE, when frac is NULL, and when the blinded data show no effect in the direction of the alternative
N2	Planned sizes of group 2 in the strata, or NULL (default)
frac	Fractions of the strata used to divide the re-estimated total sample size, see StratSampleSize. Default is NULL, which takes the fractions of the planned design
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference, 'RR' for the risk ratio or 'OR' for the odds ratio
ss.method	Re-estimation rule. 'exact' (default) re-estimates the sample size of StratSampleSize with method = 'exact', and 'standard' that of the normal approximation. With three or more strata only 'standard' is available
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
ss.strata	Blinded estimate of the re-estimation. 'stratified' (default) estimates the pooled response probability of each stratum from its own interim patients, and 'pooled' uses the proportion of responders among all the interim patients for every stratum, as a re-estimation that ignores the strata would. With 'pooled' the final group sizes depend only on the total number of interim responders

search	Search of the exact re-estimation, see StratSampleSize; it does not go beyond the largest size that N.max allows
search.limit	Limit of the search 'stable', see StratSampleSize
N.min	Smallest final total sample size, or NULL (default)
N.max	Largest final total sample size, or NULL (default)
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the group sizes. The others, 'stratum', 'group' and 'both' of StratPower, make the sizes random and need two strata, see Details
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4
ss.frac	Fractions of the strata used by the re-estimation with a random allocation: 'observed' (default), those of the interim patients, or 'planned', those given by frac

Details

The re-estimation is that of StratPowerBSSR: in each stratum the blinded estimate of the pooled response probability is divided between the groups with the assumed effect Delta.A, and the total sample size of StratSampleSize for these probabilities is divided between the strata. The final group sizes are at least the interim sizes, at least the planned sizes when restricted = TRUE, and their total lies between N.min and N.max.

With a random allocation (alloc other than 'fixed', two strata only) the re-estimation gives the final total sample size as in StratPowerBSSR, with the fractions of the strata among the interim patients (ss.frac = 'observed') or those of frac, and the remaining patients are allocated at random, so the final group sizes are NA. The attribute Power is then that of StratPower with the final total and the same allocation.

Value

An object of class bbssr_sbssr, a data frame with one row per stratum containing:

Stratum Number of the stratum

n1, n2, S Interim data

hat.p Blinded estimate of the pooled response probability

hat.p1, hat.p2 Planning probabilities of the two groups, truncated to the unit interval

N1.final, N2.final Final group sizes, NA with a random allocation

n1.stage2, n2.stage2 Numbers of patients still to be enrolled, NA with a random allocation

The attributes n.raw (unrounded total of the normal approximation), N.final (final total sample size) and Power (exact power of the final design at the planning probabilities) summarize the re-estimation, and the attribute search.limit holds the limit of the nominal total examined by search = 'stable' (otherwise NA).

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

StratPowerBSSR

Examples

```
StratBSSR(n1 = c(15, 15), n2 = c(15, 15), S = c(12, 18), Delta.A = 0.2, r = 1,
          alpha = 0.025, tar.power = 0.8, Test = 'CMH', N1 = c(30, 30),
          N2 = c(30, 30), ss.method = 'standard')
```

StratCondRejectBSSR *Conditional Rejection Probabilities of Blinded Sample Size Re-Estimation for a Stratified Design*

Description

Computes, under the null hypothesis of equal response probabilities in every stratum, the probability that a trial with two strata and blinded sample size re-estimation rejects the null hypothesis given the pooled numbers of responders of the strata at the interim analysis and in the second stage. These conditional rejection probabilities do not depend on the response probabilities of the strata, and the type I error rate is their average over the binomial distributions of the pooled counts. With `theta`, the type I error rate is also returned for each interim outcome. This is the counterpart of BinaryCondRejectBSSR for a stratified design.

Usage

```
StratCondRejectBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard"),
  ss.Test = Test,
  ss.alpha = alpha,
  ss.strata = c("stratified", "pooled"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
```

```
  n.interim = NULL,
  frac = NULL,
  theta = NULL,
  full = FALSE
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, common to the strata and used to divide the blinded estimate of the pooled response probability of each stratum between the groups at the re-estimation
N1	Planned sizes of group 1 in the strata, which are those of the fixed design
N2	Planned sizes of group 2 in the strata
omega	Fraction of the planned sizes at the interim analysis: in stratum j the interim size of group 2 is ceiling(omega N2[j]) and that of group 1 is ceiling(r[j] n2[j]). Supply either omega or n.interim
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. The blinded estimate of the pooled response probability of a stratum is divided between the groups with this ratio, also when the interim sizes are not in this ratio
alpha	Level of significance of the final analysis
tar.power	Target power of the re-estimation
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
restricted	Logical. If TRUE, the final group sizes are at least the planned sizes. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference, 'RR' for the risk ratio or 'OR' for the odds ratio
ss.method	Re-estimation rule. 'exact' (default) re-estimates the sample size of StratSampleSize with method = 'exact', and 'standard' that of the normal approximation. With three or more strata only 'standard' is available
ss.Test	Test used in the re-estimation. Default is Test

ss.alpha	Level used in the re-estimation. Default is alpha
ss.strata	Blinded estimate of the re-estimation. 'stratified' (default) estimates the pooled response probability of each stratum from its own interim patients, and 'pooled' uses the proportion of responders among all the interim patients for every stratum, as a re-estimation that ignores the strata would. With 'pooled' the final group sizes depend only on the total number of interim responders
search	Search of the exact re-estimation, see StratSampleSize; it does not go beyond the largest size that N.max allows
search.limit	Limit of the search 'stable', see StratSampleSize
N.min	Smallest final total sample size, or NULL (default)
N.max	Largest final total sample size, or NULL (default)
n.interim	Interim group sizes, as a matrix with two rows (group 1 and group 2) and one column per stratum. Supply either omega or n.interim
frac	Fractions of the strata used to divide the re-estimated total sample size between the strata, see StratSampleSize. Default is NULL, which takes the fractions of the planned design, (N1 + N2) / sum(N1 + N2)
theta	Optional common response probabilities of the strata at which the type I error rate is decomposed by the interim outcome: a vector, whose values form a grid of pairs, or a matrix with two columns that gives the points directly. Default is NULL, which skips the decomposition. The time of the decomposition grows with the number of points
full	Logical. If TRUE, the attribute full lists the conditional rejection probability of every interim and second-stage outcome. Default is FALSE. A listing of more than 1e6 rows is refused

Details

Let n1[j] and n2[j] be the interim group sizes of stratum j, and m1[j] and m2[j] the second-stage group sizes reached from the interim outcome s = (s1, s2). Under the null hypothesis both groups of stratum j respond with the probability theta_j. Given s_j, the number of interim responders of group 1 in stratum j is hypergeometric, being the number of group 1 patients among s_j responders drawn from the n1[j] + n2[j] interim patients of the stratum. Given the pooled number w_j of second-stage responders of the stratum, the number of second-stage responders of group 1 is an independent hypergeometric count of the same kind. CRP sums the rejection region of the final design over these four distributions and does not depend on theta. The type I error rate is the sum over s and w = (w1, w2) of b(s1; n1[1] + n2[1], theta_1) b(s2; n1[2] + n2[2], theta_2) b(w1; m1[1] + m2[1], theta_1) b(w2; m1[2] + m2[2], theta_2) CRP(s, w), where b is the binomial probability.

CRP.total is the conditional size of the final test given the final numbers of responders z_j = s_j + w_j of the strata, with the number of responders of group 1 of each stratum hypergeometric among all its final patients. The sum of the contributions at a point equals the type I error rate of StratTypeIErrorBSSR there.

Only designs with two strata are covered, as for the exact computations of StratTypeIErrorBSSR.

Value

An object of class bbssr_scrp, a data frame with one row for every pair of interim pooled numbers of responders of the strata, the first stratum varying fastest, containing:

s1, s2 Pooled numbers of responders of the two strata at the interim analysis

N1.S1, N1.S2, N2.S1, N2.S2 Final group sizes reached from (s1, s2)

CRP.max, CRP.min Largest and smallest rejection probability over the pooled numbers of responders (w1, w2) of the strata in the second stage

exceed Number of pairs (w1, w2) at which the rejection probability exceeds alpha

CRP.total.max Largest conditional size of the final test given the final numbers of responders of the strata, see Details

The attribute max is a data frame with the largest values of CRP and CRP.total over all outcomes, and the first outcome, in the order of the rows and then of (w1, w2) with w1 varying fastest, at which each is attained up to rounding error. The attribute exceed gives the numbers of outcomes (s1, s2, w1, w2) at which each of the two exceeds alpha, and the attribute reestimation holds the final group sizes for every interim outcome, as in StratPowerBSSR. When theta is supplied, the attribute by.s is a data frame with one row for each point and interim outcome, containing the probability prob.s of (s1, s2), the type I error rate TIE.s given (s1, s2) and their product contribution, and the attribute TIE gives the type I error rate at each point. With full = TRUE the attribute full is a data frame with one row for every outcome (s1, s2, w1, w2), ordered by s2, s1, w2 and w1, with the columns CRP and CRP.total.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

StratTypeIErrorBSSR, BinaryCondRejectBSSR

Examples

```
crp <- StratCondRejectBSSR(Delta.A = 0.3, N1 = c(8, 8), N2 = c(8, 8), omega = 0.5,
                           r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                           ss.method = 'standard', theta = c(0.3, 0.5))
print(crp)
attr(crp, 'TIE')
```

StratGridBSSR *Power and Sample Size of Several Stratified Re-estimation Designs*

Description

Evaluates a set of stratified designs with blinded sample size re-estimation over a common set of scenarios of true pooled response probabilities of the strata, by calling StratPowerBSSR for each design, and returns the results as one data frame with a row for each design and each scenario.

Usage

```
StratGridBSSR(design, p, Delta.T = NULL, type1 = FALSE, verbose = FALSE, ...)
```

Arguments

design	Data frame with one row per design. Its columns are arguments of StratPowerBSSR other than p and n.interim, for example Test, Delta.A, omega or N.max. An argument with one value per stratum is given by one column per stratum, named with the suffixes .S1, .S2, ..., for example N1.S1 and N1.S2; a column without the suffix gives a value common to the strata, where the argument allows it (r and Delta.T). The interim group sizes are given by the columns n1.interim.S1, ... and n2.interim.S1, ..., and the planned group sizes either by N1.S1, ... and N2.S1, ... or by the planning probabilities p.plan.S1, ..., see Details. A missing value in a column means that the argument is not given for that design
p	True pooled response probabilities of the strata at which every design is evaluated: a vector with one element per stratum, or a matrix with one row per scenario and one column per stratum. Every design must have this number of strata. Scenarios in which a response probability of a group falls outside the unit interval are dropped for that design, and every design needs at least one scenario that is kept
Delta.T	True treatment effect, a single value common to the strata or one value per stratum. The default of NULL uses the assumed effect Delta.A of each design. It can also be given in design, but not in both places
type1	Logical. If TRUE, the largest type I error rate of each design and of the fixed design is added, as computed by StratTypeIErrorBSSR with its default grid of theta and, with two strata, the certified maximum. With three or more strata the rates are simulated at every point of the grid, which takes long. Default is FALSE
verbose	Logical. If TRUE, a message is issued as each design is evaluated. Default is FALSE
...	Arguments shared by all designs, with the names of the arguments of StratPowerBSSR. An argument with one value per stratum is given here as a vector, for example N1 = c(20, 20), and the interim group sizes as n1.interim and n2.interim. An argument may not be given both here and in design

Details

When a design has the planning probabilities p.plan.S1, ... instead of the planned group sizes, the planning probabilities of the two groups of each stratum are obtained from them and the assumed effect Delta.A, as in the re-estimation, and the planned group sizes are those of StratSampleSize with the fractions frac, which must then be given, the allocation ratio, the test, the level and the target power of the design, the method ss.method of the re-estimation and the fixed allocation.

Designs with a random allocation (alloc) or with three or more strata use the random numbers of R in the order of the designs, so a call to set.seed() before StratGridBSSR makes the results reproducible.

Value

An object of class bbssr_sgrid, which inherits from bbssr_grid: a data frame with one row for each design and each scenario in which the response probabilities lie in the unit interval, containing the index design of the design, the columns of design (other than the group sizes and Delta.T), and:

N1.S1, ..., N2.S1, ... Planned group sizes

n1.interim.S1, ..., n2.interim.S1, ... Interim group sizes

Delta.T.S1, ... True treatment effect in the strata

scenario Row of p

p.S1, ..., p1.S1, ..., p2.S1, ... True pooled response probabilities of the strata and those of the two groups

power.BSSR, power.TRAD Power of the design with re-estimation and of the fixed design

se.BSSR, se.TRAD Monte Carlo standard errors, present when a design has them (three or more strata, or a random allocation) and NA for the other designs

E.N, SD.N Expected value and standard deviation of the final total sample size

N.q25, N.q50, N.q75 Quartiles of the final total sample size

P.N.max Probability that the final total sample size is the largest the design can reach, see summary.bbssr_spowerbssr

The attribute designs has one row per design with its index, the columns of design other than the group sizes and Delta.T, the planned and interim group sizes and the true effect and, with type1 = TRUE, the largest type I error rates TIE.BSSR and TIE.TRAD, the probabilities theta.S1.BSSR, ... and theta.S1.TRAD, ... at which they occur, the certified upper bounds bound.BSSR and bound.TRAD (NA without the certificate) and, with a random allocation, the standard errors se.BSSR and se.TRAD at the largest rates. Warnings and errors raised while a design is evaluated are prefixed with the index of the design. summary.bbssr_grid summarizes each design over the scenarios, and plot.bbssr_sgrid draws the results.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

StratPowerBSSR, StratTypeIErrorBSSR, StratSampleSize, BinaryGridBSSR

Examples

```
design <- data.frame(Test = c('CMH', 'Strat-Fisher'))
grid <- StratGridBSSR(design, p = rbind(c(0.3, 0.5), c(0.4, 0.6)), Delta.A = 0.3,
                      N1 = c(10, 10), N2 = c(10, 10), omega = 0.5, r = 1,
                      alpha = 0.05, tar.power = 0.8, ss.method = 'standard')
summary(grid)


plot(grid, colour.by = 'Test')
```

```
# Planned group sizes from the planning probabilities of the strata
design <- data.frame(Delta.A = c(0.25, 0.3), p.plan.S1 = 0.3, p.plan.S2 = 0.5)
grid <- StratGridBSSR(design, p = c(0.3, 0.5), frac = c(0.5, 0.5), omega = 0.5, r = 1,
                      alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                      ss.method = 'standard')
attr(grid, 'designs')
```

`StratPower` *Exact Power of a Stratified Test*

Description

Calculates the exact power of a test of `StratRR` for a two-arm trial with a binary endpoint, with fixed group sizes in each stratum or, with two strata, with stratum and group sizes that are random.

Usage

```
StratPower(
  p1,
  p2,
  N1 = NULL,
  N2 = NULL,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  alloc = c("fixed", "stratum", "group", "both"),
  N = NULL,
  frac = NULL,
  r = 1,
  block = 4
)
```

Arguments

`p1`	Response probabilities of group 1 in the strata: a vector with one element per stratum, or a matrix with one row per scenario and one column per stratum
`p2`	Response probabilities of group 2 in the strata, of the same form as `p1`
`N1`	Sizes of group 1 in the strata, one element per stratum, with `alloc = 'fixed'`
`N2`	Sizes of group 2 in the strata, one element per stratum, with `alloc = 'fixed'`
`alpha`	Level of significance, for the alternative specified
`Test`	Statistical test. Options: `'CMH'` (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), `'CMH-cc'` (the same with the continuity correction of `mantelhaen.test`), `'Cochran'` (Cochran's test with

	the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the group sizes N1 and N2. The others need two strata and make the sizes random, see Details: 'stratum' (stratum sizes binomial, permuted blocks within the strata), 'group' (stratum sizes fixed, group sizes binomial) and 'both' (stratum and group sizes binomial)
N	Total sample size, with a random allocation
frac	Fractions of the two strata, with a random allocation. They are rescaled to sum to 1
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum, with a random allocation. Default is 1
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4

Details

The numbers of responders of the two groups in each stratum are independent binomial variables. The power is the probability of the rejection region of StratRR, computed exactly by enumerating the numbers of responders of the strata and carrying the distribution of the partial sum of S over the strata. The response probabilities may differ between the strata in any way, so the power is available also when the odds ratio or the risk difference is not common to the strata.

With alloc other than 'fixed' the design has two strata and a total of N patients, and the sizes of the strata and of the groups within them are random. With 'group' the stratum sizes are fixed at n1 = floor(N frac[1]) and n2 = N - n1, and each patient of stratum j is in group 1 with the probability r[j] / (1 + r[j]), so that the size of group 1 is binomial (Jung, 2014, Section 3.2, where the size of stratum 1 is the round-off of N frac[1]; the integer part used here reproduces all but one of the sample sizes of the column 'n fixed' of Table 4 of the article). With 'both' the size of stratum 1 is binomial as well, with the probability frac[1] (Jung, 2014, Section 3.1). With 'stratum' the stratum sizes are binomial and the groups are allocated by permuted blocks of size block within each stratum: every complete block holds block r[j] / (1 + r[j]) patients of group 1, and the number of patients of group 1 among the patients of the last, incomplete block is hypergeometric. The power is the sum over the stratum and group sizes of their probability times the power of the fixed design they give. Configurations with a probability below 1e-14 are left out, and their total probability, which bounds the error of the power, is returned in the attribute skipped. A group without patients in a stratum leaves that stratum without variation of S, and a stratum without

patients leaves a design with one stratum. The number of configurations grows with N most quickly for 'both': at N = 100 the power of one scenario takes a fraction of a second with 'stratum', about a second with 'group' and tens of seconds with 'both'.

Value

An object of class bbssr_spower, a data frame with one row per scenario containing the response probabilities p1.S1, p1.S2, ... and p2.S1, p2.S2, ... and the exact power Power. With a random allocation the attribute skipped is the total probability of the configurations left out of the sum, see Details, and n.strata the stratum sizes of alloc = 'group'.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Jung SH (2014). Stratified Fisher's exact test and its sample size calculation. *Biometrical Journal*, 56(1), 129-140.

See Also

StratRR, StratSampleSize

Examples

```
StratPower(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), N1 = c(40, 40), N2 = c(40, 40),
           alpha = 0.025, Test = 'CMH')

# Several scenarios, and the stratified exact test
StratPower(p1 = rbind(c(0.5, 0.7), c(0.45, 0.65)), p2 = rbind(c(0.3, 0.5), c(0.3, 0.5)),
           N1 = c(40, 40), N2 = c(40, 40), alpha = 0.025, Test = 'Strat-Fisher')

# Random stratum sizes, with permuted blocks of 4 within the strata
StratPower(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), alpha = 0.025, Test = 'CMH',
           alloc = 'stratum', N = 160, frac = c(0.5, 0.5))
```

StratPowerBSSR *Power of Blinded Sample Size Re-Estimation for a Stratified Design*

Description

Calculates the power and the distribution of the final total sample size of a two-arm trial with a binary endpoint, stratified into two or more strata, with blinded sample size re-estimation, together with the power of the fixed design with the planned group sizes. The sample size is re-estimated at an interim analysis from the pooled numbers of responders of the strata, without the treatment groups, and the final analysis is a test of StratRR with all patients. With two strata the power is computed exactly, and with three or more strata by simulation.

Usage

```
StratPowerBSSR(
  p,
  Delta.A,
  Delta.T,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard"),
  ss.Test = Test,
  ss.alpha = alpha,
  ss.strata = c("stratified", "pooled"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  frac = NULL,
  nsim = 1e+05,
  alloc = c("fixed", "stratum", "group", "both"),
  block = 4,
  ss.frac = c("observed", "planned"),
  nseq = 100
)
```

Arguments

p	True pooled response probabilities of the strata: a vector with one element per stratum, or a matrix with one row per scenario and one column per stratum
Delta.A	Assumed treatment effect, on the scale given by `effect`, common to the strata and used to divide the blinded estimate of the pooled response probability of each stratum between the groups at the re-estimation
Delta.T	True treatment effect, on the scale given by `effect`: a single value common to the strata, or one value per stratum
N1	Planned sizes of group 1 in the strata, which are those of the fixed design
N2	Planned sizes of group 2 in the strata
omega	Fraction of the planned sizes at the interim analysis: in stratum `j` the interim size of group 2 is `ceiling(omega N2[j])` and that of group 1 is `ceiling(r[j] n2[j])`. Supply either `omega` or `n.interim`

r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. The blinded estimate of the pooled response probability of a stratum is divided between the groups with this ratio, also when the interim sizes are not in this ratio
alpha	Level of significance of the final analysis
tar.power	Target power of the re-estimation
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
restricted	Logical. If TRUE, the final group sizes are at least the planned sizes. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference, 'RR' for the risk ratio or 'OR' for the odds ratio
ss.method	Re-estimation rule. 'exact' (default) re-estimates the sample size of StratSampleSize with method = 'exact', and 'standard' that of the normal approximation. With three or more strata only 'standard' is available
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
ss.strata	Blinded estimate of the re-estimation. 'stratified' (default) estimates the pooled response probability of each stratum from its own interim patients, and 'pooled' uses the proportion of responders among all the interim patients for every stratum, as a re-estimation that ignores the strata would. With 'pooled' the final group sizes depend only on the total number of interim responders
search	Search of the exact re-estimation, see StratSampleSize; it does not go beyond the largest size that N.max allows
search.limit	Limit of the search 'stable', see StratSampleSize
N.min	Smallest final total sample size, or NULL (default)
N.max	Largest final total sample size, or NULL (default)
n.interim	Interim group sizes, as a matrix with two rows (group 1 and group 2) and one column per stratum. Supply either omega or n.interim

frac	Fractions of the strata used to divide the re-estimated total sample size between the strata, see StratSampleSize. Default is NULL, which takes the fractions of the planned design, (N1 + N2) / sum(N1 + N2)
nsim	Number of simulated trials for each scenario with three or more strata. Default is 1e5. Ignored with two strata
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the planned and interim group sizes. The others make the sizes random, see Details: 'stratum' (stratum sizes random, permuted blocks within the strata) and 'both' (stratum and group sizes random), with any number of strata, and 'group' (stratum sizes in the fixed fractions, group sizes random), with two strata
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4
ss.frac	Fractions of the strata used by the re-estimation with a random allocation: 'observed' (default), those of the interim patients, or 'planned', those given by frac
nseq	Number of allocation sequences drawn with a random allocation and two strata. Default is 100

Details

In stratum j the blinded estimate of the pooled response probability is the number s_j of interim responders divided by the interim size of the stratum. It is divided between the groups with the assumed effect Delta.A and the allocation ratio of the stratum, as in BinaryPowerBSSR, and the total sample size is re-estimated from these probabilities by the rule ss.method. The total is divided between the strata and the groups as in StratSampleSize, and the final group sizes are at least the interim sizes, at least the planned sizes when restricted = TRUE, with a total of at least N.min and at most N.max. When the probabilities of every stratum show no effect in the direction of the alternative, which happens with effect = 'RR' when no interim patient responds and with effect = 'OR' when in each stratum no interim patient or every interim patient responds, the planned group sizes are kept. A stratum whose pooled estimate is close to 0 with effect = 'RR', or close to 0 or 1 with effect = 'OR', has planning probabilities close to each other, so the re-estimated size can be very large, and the time of the exact computation grows with it. An upper bound N.max keeps both in check: the final total does not exceed it, and the exact search of ss.method = 'exact' does not go beyond it.

With two strata the final group sizes depend on the interim data only through (s_1, s_2). Given s_j, the number of interim responders of group 1 in stratum j follows from the two binomial distributions of the stratum, and the second-stage numbers of responders are binomial given the final sizes. The power is the sum over (s_1, s_2) of the probability of the interim outcome times the conditional probability of rejection, computed exactly by enumeration in compiled code.

With three or more strata the number of interim outcomes grows as a product over the strata, and the power is estimated from nsim simulated trials, with the random numbers of R, so that set.seed() makes the results reproducible. The power of the fixed design with the fixed allocation is exact for any number of strata.

With a random allocation (alloc other than 'fixed') and two strata the patients enter the strata and the groups as in StratPower, the interim analysis has the first sum(n1 + n2) patients of the interim sizes above, and the final analysis the first N patients, where N is the re-estimated total. The re-estimation computes the total as above, for the fixed allocation, with the fractions of the strata among

the interim patients (`ss.frac = 'observed'`) or those of `frac`, and takes the total of the group sizes it gives, within the same bounds; the allocation of the remaining patients is random. Given an allocation sequence, that is the strata and the groups of the patients in the order of enrolment, the design has fixed group sizes and its power is computed exactly as above. The power is the average over `nseq` sequences drawn with the random numbers of R, so that `set.seed()` makes the results reproducible, and `se.BSSR` is the standard deviation of the powers of the sequences divided by `sqrt(nseq)`. The power of the fixed design with the planned total and the same allocation is exact. A sequence without interim patients in a stratum is drawn again.

With a random allocation and three or more strata (`alloc = 'stratum'` or `'both'`) the allocation is simulated with the responses, in `nsim` trials. The strata of the first `sum(n1 + n2)` patients are multinomial with the probabilities `frac`, drawn again until every stratum has one of them, and their groups follow the allocation within the strata. The total is re-estimated as with two strata, and the strata and the groups of the remaining patients continue the allocation, the permuted blocks of `alloc = 'stratum'` included. The power of the fixed design with the planned total and the same allocation is estimated from `nsim` other trials, with the standard error `se.TRAD`. A group without patients is allowed in both designs, and the fixed design is analysed without a stratum that has no patients.

Value

An object of class `bbssr_spowerbssr`, a data frame with one row for each scenario in which both response probabilities of every stratum lie in the unit interval, containing:

p.S1, p.S2, ... True pooled response probabilities of the strata

p1.S1, ..., p2.S1, ... True response probabilities of the two groups

power.BSSR Power of the design with re-estimation

power.TRAD Power of the fixed design with the planned group sizes

E.N Expected final total sample size

se.BSSR Monte Carlo standard error of `power.BSSR`, with three or more strata or with a random allocation

se.TRAD Monte Carlo standard error of `power.TRAD`, with three or more strata and a random allocation

The attribute `reestimation` holds, with two strata, the final group sizes for every pair of interim pooled numbers of responders, `N.dist` the distribution of the final total sample size in each scenario, `scenario` the rows of `p` kept as scenarios, and the other attributes the design settings. `summary.bbssr_spowerbssr` summarizes the distribution of the final total.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Woolson RF, Bean JA, Rojas PB (1986). Sample size for case-control studies using Cochran's statistic. *Biometrics*, 42(4), 927-932.

Nam J (1992). Sample size determination for case-control studies and the comparison of stratified and unstratified analyses. *Biometrics*, 48(2), 389-395.

Jung SH (2014). Stratified Fisher's exact test and its sample size calculation. *Biometrical Journal*, 56(1), 129-140.

See Also

StratGridBSSR, StratTypeIErrorBSSR, StratBSSR

Examples

```
res <- StratPowerBSSR(p = rbind(c(0.3, 0.5), c(0.4, 0.6)), Delta.A = 0.2,
                      Delta.T = 0.2, N1 = c(30, 30), N2 = c(30, 30), omega = 0.5,
                      r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                      ss.method = 'standard')
print(res)
```

StratRR *Rejection Region of a Stratified Test*

Description

Returns the critical values of a test for a two-arm trial with a binary endpoint stratified into two or more strata, for every combination of the numbers of responders of the strata. The tests are based on S, the number of responders of group 1 summed over the strata. Given the numbers of responders z = (z_1, ..., z_J) of the strata, each test rejects the null hypothesis when S is at most a lower or at least an upper critical value. The unstratified tests, which ignore the strata, are included for comparison.

Usage

```
StratRR(
  N1,
  N2,
  alpha,
  Test,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker")
)
```

Arguments

N1	Sizes of group 1 in the strata, one element per stratum
N2	Sizes of group 2 in the strata, one element per stratum
alpha	Level of significance, for the alternative specified

`Test`	Statistical test. Options: `'CMH'` (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), `'CMH-cc'` (the same with the continuity correction of `mantelhaen.test`), `'Cochran'` (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), `'Cochran-cc'` (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), `'Strat-Fisher'` (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: `'Unstrat-Chisq'` (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), `'Unstrat-Chisq-cc'` (the same with the continuity correction) and `'Unstrat-Fisher'` (Fisher's exact test). See Details
`alternative`	Direction of the alternative hypothesis. Options: `'greater'` (default), under which group 1 responds more often than group 2, `'less'` or `'two.sided'`
`tsmethod`	Convention of the two-sided p-value of `'Strat-Fisher'` and `'Unstrat-Fisher'`, as for the Fisher exact test of `BinaryRR`. Options: `'minlike'` (default), `'central'` or `'blaker'`. Ignored by the other tests and for one-sided alternatives

Details

In stratum `j`, with `N_j = N1[j] + N2[j]` patients and `z_j` responders, the null expectation of the number of responders of group 1 is `N1[j] z_j / N_j`, and `E` is the sum of these over the strata. The Cochran-Mantel-Haenszel statistic is `(S - E) / sqrt(V)` with the hypergeometric variance `V = sum_j N1[j] N2[j] z_j (N_j - z_j) / (N_j^2 (N_j - 1))`, and Cochran's statistic uses the binomial variance `sum_j N1[j] N2[j] z_j (N_j - z_j) / N_j^3` instead. The continuity correction reduces `|S - E|` by 0.5 when it is at least 0.5, as in `mantelhaen.test`; a deviation of exactly 0.5 is recognised although `E`, a sum over the strata, carries rounding errors. Nam (1992) reduces `S - E` by 0.5 in every outcome, which makes a difference only where `|S - E| < 0.5`. There `S` is the integer next to `E`, and since the hypergeometric variance is at least `f (1 - f)`, with `f` the fractional part of `E`, and the binomial variance at least half of it, the one-sided p-value of `'Cochran-cc'` exceeds 0.078. The two corrections therefore reject the same outcomes at one-sided levels up to 0.078. The statistics are referred to the standard normal distribution, with the absolute value for a two-sided alternative, and an outcome whose variance is 0 is not rejected. The stratified exact conditional test uses the distribution of `S` given `z`, the convolution of the hypergeometric distributions of the strata. The two-sided p-value `'central'` is twice the smaller tail probability (Jung, 2014), and `'minlike'` and `'blaker'` order the values of `S` by their probability and by their smaller tail probability, with the ties of `BinaryRR`. As in the other functions of the package, a p-value is rejected when it is below `alpha` by more than about 1.49e-8.

The unstratified tests apply the corresponding test of a single 2 x 2 table to the table of totals, with `sum_j N1[j]` and `sum_j N2[j]` patients and `sum_j z_j` responders, so their critical values depend on `z` only through this total. Pearson's statistic of that table is Cochran's statistic of one stratum, as in `'Chisq'` of `BinaryRR`, and the continuity correction is that of `'Cochran-cc'`; with the correction applied to every outcome, as in the z_u test of Nam (1992, Section 2.2), the same outcomes are rejected at one-sided levels up to 0.078. An unstratified test is valid when the distribution of the strata is the same in the two groups, and its level can be distorted otherwise (Nam, 1992; Jung, 2014); its power and type I error rate are computed here with the response probabilities of each stratum, so these consequences can be evaluated.

The function stores the critical values for reuse during the session, like the rejection regions of the other designs.

Value

An object of class `bbssr_srr`, a data frame with one row for each combination of the numbers of responders of the strata, the first stratum varying fastest, containing:

z.S1, z.S2, ... Numbers of responders of the strata

S.min, S.max Smallest and largest possible value of `S`

lower Lower critical value: the null hypothesis is rejected when `S <= lower`. `NA` when no value of `S` is rejected in the lower tail

upper Upper critical value: the null hypothesis is rejected when `S >= upper`. `NA` when no value of `S` is rejected in the upper tail

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Woolson RF, Bean JA, Rojas PB (1986). Sample size for case-control studies using Cochran's statistic. *Biometrics*, 42(4), 927-932.

Nam J (1992). Sample size determination for case-control studies and the comparison of stratified and unstratified analyses. *Biometrics*, 48(2), 389-395.

Jung SH (2014). Stratified Fisher's exact test and its sample size calculation. *Biometrical Journal*, 56(1), 129-140.

See Also

`StratPower`, `StratSampleSize`

Examples

```
rr <- StratRR(N1 = c(10, 12), N2 = c(10, 12), alpha = 0.025, Test = 'CMH')
print(rr)
```

`StratSampleSize` *Sample Size of a Stratified Test*

Description

Calculates the total sample size of a two-arm trial with a binary endpoint stratified into two or more strata, for a test of `StratRR`, either exactly or by the normal approximation of Woolson et al. (1986) and Nam (1992).

Usage

```
StratSampleSize(
  p1,
  p2,
  frac,
  r = 1,
  alpha,
  tar.power,
  Test,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  method = c("exact", "standard"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  alloc = c("fixed", "stratum", "group", "both"),
  block = 4
)
```

Arguments

p1	Response probabilities of group 1 in the strata, one element per stratum
p2	Response probabilities of group 2 in the strata, one element per stratum
frac	Fractions of the strata, one positive element per stratum. They are rescaled to sum to 1, so they may also be given as numbers of patients
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. Default is 1
alpha	Level of significance, for the alternative specified
tar.power	Target power
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
method	Calculation of the sample size. 'exact' (default) searches for the total sample size at which the exact power of StratPower attains tar.power. 'standard' uses the normal approximation, see Details

search	How the exact calculation chooses among the sizes at which the power fluctuates around tar.power, as in BinarySampleSize. Options: 'crossing' (default), 'smallest' or 'stable'
search.limit	Limit of the search 'stable': the nominal sizes are examined downwards from max(ceiling(search.limit[1] n), ceiling(n + search.limit[2])), where n is the nominal size of the normal approximation. Default is c(2, 50)
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the group sizes N1 and N2. The others need two strata and make the sizes random, see Details: 'stratum' (stratum sizes binomial, permuted blocks within the strata), 'group' (stratum sizes fixed, group sizes binomial) and 'both' (stratum and group sizes binomial)
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4

Details

A design is described by the fractions frac of the strata and the allocation ratios r. For a nominal total sample size n, the size of group 2 in stratum j is ceiling(frac[j] n / (1 + r[j])) and that of group 1 is ceiling(r[j] N2[j]), so the total N is at least n. The exact calculation searches over n within this family of designs.

The normal approximation is that of Cochran's test. With s[j] = r[j] / (1 + r[j]), w[j] = frac[j] s[j] (1 - s[j]), D = sum_j w[j] (p1[j] - p2[j]), v0 = sum_j w[j] pbar[j] (1 - pbar[j]) with pbar[j] = s[j] p1[j] + (1 - s[j]) p2[j], and v1 = sum_j w[j] ((1 - s[j]) p1[j] (1 - p1[j]) + s[j] p2[j] (1 - p2[j])), the total sample size is N* = (z.a sqrt(v0) + z.b sqrt(v1))^2 / D^2, formula (2.8) of Woolson et al. (1986), where z.a is the upper alpha quantile (alpha / 2 for a two-sided alternative) and z.b the tar.power quantile of the standard normal distribution. For the tests with a continuity correction it is N = (1 + sqrt(1 + 2 / (D N*)))^2 N* / 4, formula (2.3) of Nam (1992). The same approximation is used for every test, since the Cochran-Mantel-Haenszel statistic differs from Cochran's only by the factors (N_j - 1) / N_j of the variance, and it serves as the starting value of the exact search. The nominal total is ceiling(N*), or ceiling(N); the published tables of Woolson et al. (1986) and Nam (1992) compute N* and N with the quantiles rounded to three decimals, such as 1.645 for the level 0.05 and 0.841 or 0.842 for the power 0.8, and round them to the nearest integer instead.

For the unstratified tests the approximation is that of the comparison of two proportions on the table of totals, formula (3.3) of Woolson et al. (1986), or (2.3) of Nam (1992) with one stratum for 'Unstrat-Chisq-cc': the formula above with one stratum, in which group 1 holds the fraction sum_j frac[j] s[j] of the patients and the response probability of each group is the average of its probabilities over the strata, weighted by its sizes in the strata. The exact power of these tests is that of the stratified design, with the response probabilities of each stratum, so the search accounts for the difference between the strata that the table of totals ignores.

The exact power is not monotone in the sample size, so the result depends on search, as described in BinarySampleSize.

With a random allocation (alloc other than 'fixed', two strata only) the total sample size N is searched directly, the nominal total being the total itself, and the power is that of StratPower with the same allocation. Jung (2014) searches upwards from the sample size of the Mantel-Haenszel test. The power is not monotone, so the result can differ from that of search = 'crossing', which starts from the normal approximation.

Value

An object of class bbssr_ssamplesize, a list containing the arguments p1, p2, frac, r, alpha, tar.power, Test and alternative, the group sizes N1 and N2 of the strata (NULL with a random allocation), their total N, and the exact power Power of the design. The attribute n.raw holds the unrounded total of the normal approximation, n the nominal total of the design and search.limit the limit of the nominal total examined by search = 'stable' (otherwise NA). With a random allocation the attributes skipped and n.strata are those of StratPower.

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

References

Woolson RF, Bean JA, Rojas PB (1986). Sample size for case-control studies using Cochran's statistic. *Biometrics*, 42(4), 927-932.

Nam J (1992). Sample size determination for case-control studies and the comparison of stratified and unstratified analyses. *Biometrics*, 48(2), 389-395.

Jung SH (2014). Stratified Fisher's exact test and its sample size calculation. *Biometrical Journal*, 56(1), 129-140.

See Also

StratPower, StratPowerBSSR

Examples

```
# Normal approximation and exact sample size with two strata
StratSampleSize(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), frac = c(0.5, 0.5),
                alpha = 0.025, tar.power = 0.8, Test = 'CMH', method = 'standard')
StratSampleSize(p1 = c(0.5, 0.7), p2 = c(0.3, 0.5), frac = c(0.5, 0.5),
                alpha = 0.025, tar.power = 0.8, Test = 'CMH')
```

StratTypeIErrorBSSR *Type I Error Rate of Blinded Sample Size Re-Estimation for a Stratified Design*

Description

Calculates the type I error rate of a two-arm trial with a binary endpoint, stratified into two or more strata, with blinded sample size re-estimation, over the common response probabilities of the strata under the null hypothesis, together with that of the fixed design with the planned group sizes. With two strata the rate is computed exactly and its largest value over a rectangle of probabilities is certified; with three or more strata the rate of the re-estimation design is estimated by simulation.

Usage

```
StratTypeIErrorBSSR(
  Delta.A,
  N1,
  N2,
  omega = NULL,
  r,
  alpha,
  tar.power,
  Test,
  restricted = FALSE,
  alternative = c("greater", "less", "two.sided"),
  tsmethod = c("minlike", "central", "blaker"),
  effect = c("RD", "RR", "OR"),
  ss.method = c("exact", "standard"),
  ss.Test = Test,
  ss.alpha = alpha,
  ss.strata = c("stratified", "pooled"),
  search = c("crossing", "smallest", "stable"),
  search.limit = c(2, 50),
  N.min = NULL,
  N.max = NULL,
  n.interim = NULL,
  frac = NULL,
  theta = seq(0, 1, by = 0.05),
  maximize = c("certified", "grid"),
  nsim = 1e+05,
  alloc = c("fixed", "stratum", "group", "both"),
  block = 4,
  ss.frac = c("observed", "planned"),
  nseq = 100
)
```

Arguments

Delta.A	Assumed treatment effect, on the scale given by effect, common to the strata and used to divide the blinded estimate of the pooled response probability of each stratum between the groups at the re-estimation
N1	Planned sizes of group 1 in the strata, which are those of the fixed design
N2	Planned sizes of group 2 in the strata
omega	Fraction of the planned sizes at the interim analysis: in stratum j the interim size of group 2 is ceiling(omega N2[j]) and that of group 1 is ceiling(r[j] n2[j]). Supply either omega or n.interim
r	Allocation ratio of group 1 to group 2 within the strata, a single value or one value per stratum. The blinded estimate of the pooled response probability of a stratum is divided between the groups with this ratio, also when the interim sizes are not in this ratio

alpha	Level of significance of the final analysis
tar.power	Target power of the re-estimation
Test	Statistical test. Options: 'CMH' (Cochran-Mantel-Haenszel test with the hypergeometric variance, without continuity correction), 'CMH-cc' (the same with the continuity correction of mantelhaen.test), 'Cochran' (Cochran's test with the binomial variance, the test of Woolson et al. (1986)), 'Cochran-cc' (the same with the continuity correction, which rejects the same outcomes as that of Nam (1992) at one-sided levels up to 0.078), 'Strat-Fisher' (the stratified exact conditional test of Jung (2014)), or one of the unstratified tests of the table of totals: 'Unstrat-Chisq' (Pearson's chi-squared test, the comparison of two proportions with the pooled variance), 'Unstrat-Chisq-cc' (the same with the continuity correction) and 'Unstrat-Fisher' (Fisher's exact test). See Details
restricted	Logical. If TRUE, the final group sizes are at least the planned sizes. Default is FALSE
alternative	Direction of the alternative hypothesis. Options: 'greater' (default), under which group 1 responds more often than group 2, 'less' or 'two.sided'
tsmethod	Convention of the two-sided p-value of 'Strat-Fisher' and 'Unstrat-Fisher', as for the Fisher exact test of BinaryRR. Options: 'minlike' (default), 'central' or 'blaker'. Ignored by the other tests and for one-sided alternatives
effect	Scale of Delta.A and Delta.T. Options: 'RD' (default) for the risk difference, 'RR' for the risk ratio or 'OR' for the odds ratio
ss.method	Re-estimation rule. 'exact' (default) re-estimates the sample size of StratSampleSize with method = 'exact', and 'standard' that of the normal approximation. With three or more strata only 'standard' is available
ss.Test	Test used in the re-estimation. Default is Test
ss.alpha	Level used in the re-estimation. Default is alpha
ss.strata	Blinded estimate of the re-estimation. 'stratified' (default) estimates the pooled response probability of each stratum from its own interim patients, and 'pooled' uses the proportion of responders among all the interim patients for every stratum, as a re-estimation that ignores the strata would. With 'pooled' the final group sizes depend only on the total number of interim responders
search	Search of the exact re-estimation, see StratSampleSize; it does not go beyond the largest size that N.max allows
search.limit	Limit of the search 'stable', see StratSampleSize
N.min	Smallest final total sample size, or NULL (default)
N.max	Largest final total sample size, or NULL (default)
n.interim	Interim group sizes, as a matrix with two rows (group 1 and group 2) and one column per stratum. Supply either omega or n.interim
frac	Fractions of the strata used to divide the re-estimated total sample size between the strata, see StratSampleSize. Default is NULL, which takes the fractions of the planned design, (N1 + N2) / sum(N1 + N2)
theta	Common response probabilities of the strata under the null hypothesis. A vector gives, with two strata, the grid of all pairs of its values, and with three or more strata the points at which every stratum has the same value. A matrix with one column per stratum gives the points directly. Default is seq(0, 1, by = 0.05)

maximize	How the largest type I error rate is located. With 'certified' (the default with two strata, and available only with two) it is certified over the rectangle from the smallest to the largest value of theta in each stratum, see Details. With 'grid' (the default with three or more strata) it is the largest value at the points of theta
nsim	Number of simulated trials at each point with three or more strata. Default is 1e5. Ignored with two strata
alloc	Allocation of the patients to the strata and the groups. 'fixed' (default) uses the planned and interim group sizes. The others make the sizes random, see Details: 'stratum' (stratum sizes random, permuted blocks within the strata) and 'both' (stratum and group sizes random), with any number of strata, and 'group' (stratum sizes in the fixed fractions, group sizes random), with two strata
block	Size of the permuted blocks of alloc = 'stratum'. Each block holds block r / (1 + r) patients of group 1, which must be a whole number. Default is 4
ss.frac	Fractions of the strata used by the re-estimation with a random allocation: 'observed' (default), those of the interim patients, or 'planned', those given by frac
nseq	Number of allocation sequences drawn with a random allocation and two strata. Default is 100

Details

Under the null hypothesis both groups of stratum j respond with the probability theta_j. Given the interim and the second-stage numbers of responders of a stratum, the numbers of responders of group 1 are hypergeometric whatever theta_j, so with two strata the type I error rate is a polynomial in (theta_1, theta_2). Its coefficients in the products of the Bernstein bases of the two strata are computed exactly, and the rate is evaluated at the points of theta from them. Over a rectangle the polynomial lies between the smallest and the largest of its coefficients there, so with maximize = 'certified' the rectangle is halved along its longer side until the largest coefficient of every piece exceeds the largest value found by at most 1e-12. The largest value is reported with this upper bound, which the type I error rate cannot exceed anywhere in the rectangle. The fixed design is certified in the same way.

With three or more strata the type I error rate of the re-estimation design is estimated at each point from nsim simulated trials, with the random numbers of R, and that of the fixed design with the fixed allocation is computed exactly.

With a random allocation and two strata the allocation sequences are drawn as in StratPowerBSSR. Given a sequence the type I error rate is a polynomial as above, and the rate of the re-estimation design is the average of these polynomials over nseq sequences, itself a polynomial, whose largest value is certified as above. The certificate holds for this average, and se.BSSR gives the Monte Carlo standard error of the average at each point, from the variation between the sequences; the attribute max gives it at the point of theta with the largest rate. The rate of the fixed design with the planned total and the same allocation is exact: the sum over the stratum and group sizes of their probability times the polynomial of the fixed design they give.

With a random allocation and three or more strata (alloc = 'stratum' or 'both') the allocation is simulated with the responses as in StratPowerBSSR, and the rates of both designs are estimated at each point from nsim trials each.

Value

An object of class bbssr_stie, a data frame with one row per point, containing the probabilities theta.S1, theta.S2, ..., the type I error rates TIE.BSSR and TIE.TRAD of the designs with re-estimation and without it, and, with three or more strata or a random allocation, the Monte Carlo standard error se.BSSR, and with three or more strata and a random allocation se.TRAD for TIE.TRAD. The attribute max is a data frame with the location, the value and, with maximize = 'certified', the upper bound bound of the largest rate of each design, and with a random allocation the standard error se of the rate at the point of theta with the largest rate (NA for the exact rate of the fixed design with two strata).

Author(s)

Gosuke Homma (<my.name.is.gosuke@gmail.com>)

See Also

StratAlphaAdjBSSR, StratPowerBSSR

Examples

```
tie <- StratTypeIErrorBSSR(Delta.A = 0.2, N1 = c(20, 20), N2 = c(20, 20), omega = 0.5,
                           r = 1, alpha = 0.025, tar.power = 0.8, Test = 'CMH',
                           ss.method = 'standard', theta = seq(0.1, 0.9, by = 0.2))
attr(tie, 'max')
```

---

summary.bbssr_grid *Summary Method for bbssr_grid Objects*

Description

Summarizes each design of BinaryGridBSSR over the true pooled response probabilities at which it was evaluated.

Usage

```
## S3 method for class 'bbssr_grid'
summary(object, ...)
```

Arguments

object	An object of class bbssr_grid
...	Further arguments, ignored

Value

A data frame with one row per design, containing the attribute `designs` of `object` followed by the smallest and the largest power of the BSSR design (`power.BSSR.min`, `power.BSSR.max`) and of the fixed-sample design (`power.TRAD.min`, `power.TRAD.max`), and the largest expected total sample size `E.N.max`

Examples

```
design <- data.frame(Test = c('Chisq', 'Fisher'))
grid <- BinaryGridBSSR(design, p = c(0.3, 0.4, 0.5), Delta.A = 0.3, N1 = 30, N2 = 30,
                       omega = 0.5, r = 1, alpha = 0.025, tar.power = 0.8,
                       ss.method = 'standard')
summary(grid)
```

---

summary.bbssr_powerbssr

Summary Method for bbssr_powerbssr Objects

Description

Summarizes the distribution of the final total sample size of a design with blinded sample size re-estimation for every true pooled response probability.

Usage

```
## S3 method for class 'bbssr_powerbssr'
summary(object, probs = c(0.25, 0.5, 0.75), ...)
```

Arguments

`object`	An object of class `bbssr_powerbssr`
`probs`	Probabilities of the reported quantiles. Default is `c(0.25, 0.5, 0.75)`
`...`	Further arguments, ignored

Details

The quantile for the probability `q` is the smallest final total sample size whose cumulative probability is at least `q`.

Value

A data frame with one row per element of `p` containing `p1`, `p2`, `p`, the expected final total sample size `E.N`, its standard deviation `SD.N`, the quantiles of the final total sample size in columns named after `probs` (for example `N.q50` for the median), and the probability `P.N.max` that the final total reaches its largest attainable value

Examples

```
res <- BinaryPowerBSSR(
  p = c(0.3, 0.45),
  Delta.A = 0.3, Delta.T = 0.3,
  N1 = 10, N2 = 10, omega = 0.5, r = 1,
  alpha = 0.025, tar.power = 0.8, Test = 'Chisq'
)
summary(res)
```

`summary.bbssr_spowerbssr` *Summary Method for bbssr_spowerbssr Objects*

Description

Summarizes the distribution of the final total sample size of a stratified design with blinded sample size re-estimation in each scenario of `StratPowerBSSR`.

Usage

```
## S3 method for class 'bbssr_spowerbssr'
summary(object, probs = c(0.25, 0.5, 0.75), ...)
```

Arguments

`object`	An object of class `bbssr_spowerbssr`
`probs`	Probabilities of the reported quantiles. Default is `c(0.25, 0.5, 0.75)`
`...`	Further arguments, ignored

Details

The distribution is the attribute `N.dist` of `object`: exact with two strata and the fixed allocation, the average over the allocation sequences with two strata and a random allocation, and that of the simulated trials with three or more strata.

Value

A data frame with one row per scenario containing the true pooled response probabilities `p.S1`, `p.S2`, `...` of the strata, those of the two groups `p1.S1`, `...` and `p2.S1`, `...`, the expected value `E.N` and the standard deviation `SD.N` of the final total sample size, its quantiles (`N.q25`, `N.q50` and `N.q75` by default, the smallest total whose distribution function reaches the probability), and the probability `P.N.max` that it is the largest total the design can reach. With two strata and the fixed allocation this is the largest final total of the re-estimation over all interim outcomes. Otherwise it is `N.max`, and `P.N.max` is `NA` when `N.max` was not given.

Examples

```
res <- StratPowerBSSR(p = c(0.4, 0.6), Delta.A = 0.3, Delta.T = 0.3, N1 = c(10, 10),
                      N2 = c(10, 10), omega = 0.5, r = 1, alpha = 0.05,
                      tar.power = 0.8, Test = 'CMH', ss.method = 'standard')
summary(res)
```

Index

BinaryAlphaAdjBSSR, 4, *25*, *40*, *42*, *93*
BinaryBSSR, 9, *25*, *76*
BinaryCondRejectBSSR, 13, *98*, *101*
BinaryGridBSSR, 17, *61*, *103*, *120*
BinaryPower, 19, *63*, *78*
BinaryPowerBSSR, *10–12*, *16–18*, 21, *38*, *40*, *64*, *65*, *79*, *109*
BinaryRR, *5–7*, *10, 11*, *14, 15*, *19–21*, *23, 24*, 26, *28*, *31–33*, *37, 38*, *53*, *68*, *83*, *92*, *96*, *99*, *105*, *108*, *112*, *114*, *118*
BinarySampleSize, *5*, *10–12*, *14, 15*, *17, 18*, *23*, 31, *37*, *56*, *70*, *84*, *115*
BinaryTypeIErrorBSSR, *6–9*, *15–18*, *24, 25*, 35, *59*

coord_cartesian, *63*, *65*, *70*

mantelhaen.test, *92*, *96*, *99*, *104*, *108*, *112*, *114*, *118*

PairedAlphaAdjBSSR, 40
PairedBSSR, 44, *78*
PairedPower, 46, *80*
PairedPowerBSSR, *45*, 48, *80*
PairedRR, *47*, 51, *66*, *81*
PairedSampleSize, *41, 42*, *45*, *49*, 54, *58*, *82*
PairedTypeIErrorBSSR, *42, 43*, *50*, 57, *83*
plot.bbssr_crp, 60
plot.bbssr_grid, 61, *71*
plot.bbssr_power, 62
plot.bbssr_powerbssr, 64, *72*
plot.bbssr_prr, 65
plot.bbssr_ptie, 67
plot.bbssr_rr, 68
plot.bbssr_samplesize, 69
plot.bbssr_sgrid, 71, *103*
plot.bbssr_spowerbssr, 72
plot.bbssr_stie, 73
plot.bbssr_tie, 74
print.bbssr_alphaadj, 75
print.bbssr_bssr, 76
print.bbssr_crp, 77
print.bbssr_pbssr, 77
print.bbssr_power, 78
print.bbssr_powerbssr, 79
print.bbssr_ppower, 79
print.bbssr_ppowerbssr, 80
print.bbssr_prr, *66*, 81
print.bbssr_psamplesize, 82
print.bbssr_ptie, 82
print.bbssr_rr, *69*, 83
print.bbssr_samplesize, 84
print.bbssr_sbssr, 84
print.bbssr_scrp, 85
print.bbssr_spower, 86
print.bbssr_spowerbssr, 87
print.bbssr_srr, 88
print.bbssr_ssamplesize, 88
print.bbssr_stie, 89
print.bbssr_tie, 90

StratAlphaAdjBSSR, 91, *120*
StratBSSR, *84*, 95, *111*
StratCondRejectBSSR, *85*, 98
StratGridBSSR, *71*, 101, *111*
StratPower, *86*, *97*, 104, *109*, *113–116*
StratPowerBSSR, *72*, *87*, *97, 98*, *101–103*, 106, *116*, *119, 120*, *122*
StratRR, *88*, *104–106*, 111, *113*
StratSampleSize, *88*, *92, 93*, *96, 97*, *99, 100*, *102, 103*, *106*, *108, 109*, 113, *113*, *118*
StratTypeIErrorBSSR, *89*, *93, 94*, *100–103*, *111*, 116
summary.bbssr_grid, *103*, 120
summary.bbssr_powerbssr, *18*, 121
summary.bbssr_spowerbssr, *103*, *110*, 122