An introduction to planning and analyzing three-arm trials using the package ThreeArmedTrials

Tobias Mütze

2022-12-16

The package `ThreeArmedTrials` provides a collection of functions for statistical inference in three-arm trials with the *gold standard* design. For a study assessing non-inferiority or superiority of an experimental treatment compared to an active control, tools to

- determine the optimal sample size allocation
- plan the sample size
- analyze the final data set

are provided for a variety of models (Poisson, negative binomial, normal, censored exponential, binary, non-parametric). This document provides an overview of the implemented statistical models and demonstrates the package's usage in worked-out examples for Poisson, binary, and the non-parametric model.

Introduction

The *gold standard* design for three-arm trials refers to trials with an active control (often called reference) and a placebo control in addition to the experimental treatment group. This trial design is recommended when being ethically justifiable and it allows the simultaneous comparison of experimental treatment, reference, and placebo. With λ_E, λ_R, and λ_P the parameter of interest for the experimental treatment, the reference, and the placebo, respectively, and smaller parameters μ_k being desired, the null hypothesis of non-inferiority and superiority can be defined by

$$H_0 : \lambda_P - \lambda_E \leq \Delta(\lambda_P - \lambda_R) \quad \text{vs.} \quad H_1 : \lambda_P - \lambda_E > \Delta(\lambda_P - \lambda_R).$$

Superiority is tested with a margin of $\Delta \in [1, \infty)$ and non-inferiority with a margin of $\Delta \in (0, 1)$. In the case of non-inferiority, this hypothesis is referred to as the retention of effect (RET) hypothesis. If the null hypothesis H_0 is rejected, non-inferiority or superiority of the experimental treatment compared to the reference is shown.

Statistical model, hypothesis, and test statistic

The outcomes under experimental treatment (E), reference (R), and placebo (P) are modeled as the random variables

$$X_{k,i} \quad i = 1, \ldots, n_k, \quad k = E, R, P.$$

The random variables are distributed according to a parametric family with densities

$$\{f(\theta, \cdot) : \theta \in \Theta\}, \Theta \subseteq \mathbb{R}^d.$$

We denote the parameter vectors of the experimental, reference, and placebo group as $\theta_E, \theta_R, \theta_P \in \Theta$, respectively.
Let $h : \Theta^d \to \mathbb{R}$ be a monotonic function and $\lambda_k := h(\theta_k), k = E, R, P$, be the parameters of interest. The next table provides an overview of the implemented models, the respective literature, and common choices of h.

Model	Literature	Distribution of $X_{k,i}$	Parameter vector θ_k	h
Normal - homoscedastic	Pigeot et al. (2003)	$\mathcal{N}(\mu_k, \sigma^2)$	(μ_k, σ^2)	$(\mu_k, \sigma^2) \mapsto \mu_k$
Binary	Kieser and Friede (2007)	$\mathcal{B}(1, p_k)$	p_k	$id, logit$
Normal - heteroscedastic	Hasler et al. (2008)	$\mathcal{N}(\mu_k, \sigma_k^2)$	(μ_k, σ_k^2)	$(\mu_k, \sigma_k^2) \mapsto \mu_k$
Poisson	Mielke and Munk (2009)	$\text{Pois}(\mu_k)$	μ_k	id
Censored exponential	Mielke et al. (2009)	$\min\left(\text{Exp}(\mu_k), U_{k,i}\right)$	(μ_k, p_k)	$(\mu_k, p_k) \mapsto \log(\mu_k)$
Negative binomial	Mütze et al. (2016a)	$\text{NB}(\mu_k, \phi)$	(μ_k, ϕ)	$(\mu_k, \phi) \mapsto \mu_k$
Non-parametric	Mütze et al. (2016b)	-	(μ_k, σ_k^2)	$(\mu_k, \sigma_k^2) \mapsto \mu_k$

For the censored exponential model, $U_{k,i}$ are the censoring times for which no particular distribution is assumed and p_k are the probabilities of an observation being uncensored. The non-parametric model does not assume any particular distribution for the data and μ_k and σ_k^2 are the expected value and the variance, respectively.

The hypothesis H_0 is rejected at significance level α if the Wald-type test statistic

$$T = \sqrt{n}\,\frac{\hat{\lambda}_E - \Delta\hat{\lambda}_R + (\Delta - 1)\hat{\lambda}_P}{\hat{\sigma}}.$$

is smaller than the α-quantile of a normal distribution (binary, censored exponential, negative binomial, and Poisson model), t-distribution (normal models), or a permutation distribution (non-parametric model). For the normal and the non-parametric models the variance estimator $\hat{\sigma}^2$ is based on group specific sample variances. For the censored exponential model the variance is estimated by an unrestricted maximum-likelihood estimator. For the binary, Poisson, and negative binomial model, the user can choose between an unrestricted maximum-likelihood variance estimator and a maximum-likelihood variance estimator restricted to the parameter space of the null hypothesis. The estimators $\hat{\lambda}_k$ are maximum-likelihood estimators for the parametric models and group specific means for the non-parametric model.

The main functions in *ThreeArmedTrials*

This package is implemented such that the function names for each of the three tasks (optimal sample size allocation calculation, sample size calculation, data analysis) do not depend on the statistical model. The model is then specified when calling the functions. The next table lists the task and the corresponding R-function for performing this task.

Task	R-function
Optimal sample size allocation calculation	`opt_alloc_RET()`
Sample size and power related calculation	`power_RET()`
Testing the retention of effect hypothesis	`test_RET()`

The functions `opt_alloc_RET()` and `power_RET()` require the user to specify a parameter vector. For each of the models, this parameter vector corresponds to the vector θ_k defined above.

Planning and analyzing a study with the binary data

In this section, we exercise planning and analyzing a three-arm study in the *gold standard* design for binary data. The example is motivated by clinical trials in patients with depression. We define the binary endpoint *remission* as a HAM-D total score of less than or equal to 7. For more details see Kieser and Friede (2007). The parameter of interest p_k is then the probability that remission is achieved. In contrast to the definition of the non-inferiority hypothesis above, larger values of p_k are more desirable. Through the choice of the function h, the parameter p_k can be transformed such that larger values of the parameter are more desirable. From a practical point of view, the appropriate choices for h are the identity function and the logit-function. Considering that the hypothesis above is defined for the case where smaller values are desired, we obtain the functions

$$h(x) = -x,$$
$$h(x) = -\log\left(\frac{x}{1-x}\right).$$

With $\lambda_k = h(p_k)$, we obtain the hypotheses

$$H_{0,id} : p_E - p_P \leq \Delta(p_R - p_P)$$
$$H_{0,logit} : \log\left(\frac{p_E}{1-p_E}\right) - \log\left(\frac{p_P}{1-p_P}\right) \leq \Delta\left(\log\left(\frac{p_R}{1-p_R}\right) - \log\left(\frac{p_P}{1-p_P}\right)\right)$$

In the following, we will focus on planning and analyzing a three-arm trial with the hypothesis $H_{0,logit}$.

Designing a trial

The trial will be designed in two steps. Firstly, we calculate an optimal sample size allocation for a given alternative. Then, we calculate the sample sample of the trial.

The following information is required for calculating the optimal sample size of a trial with the function `opt_alloc_RET()`:

- The vector of probabilities (p_E, p_R, p_P) in the alternative. Here: $(p_E, p_R, p_P) = (0.5, 0.3, 0.2)$.
- A non-inferiority margin Δ. Here: $\Delta = 0.8$.
- The function h which is used to define the hypothesis. Here, the function h is defined as `h = function(x){-log(x/(1-x))}`.

The optimal sample size allocation is calculated and given by

```
w <- ThreeArmedTrials::opt_alloc_RET(experiment = 0.5,
                                     reference = 0.3,
                                     placebo = 0.2,
                                     Delta = 0.8,
                                     distribution = "binary",
                                     h = function(x){-log(x/(1-x))})
```

The function `opt_alloc_RET()` returns a vector (w_E^*, w_R^*, w_P^*) with the optimal sample size allocation.

```
print(w)
```

```
## [1] 0.4710601 0.4111749 0.1177650
```

For calculating the sample size of a trial with the function `power_RET()`, we require the following information:

- The vector of probabilities (p_E, p_R, p_P) in the alternative. Here: $(p_E, p_R, p_P) = (0.5, 0.3, 0.2)$.
- A non-inferiority margin Δ. Here: $\Delta = 0.8$.
- A significance level and a power. Here: $\alpha = 0.025$ and $\text{power} = 0.8$.
- A sample size allocation. Here: the optimal allocation `w`
- Whether the variance in the Wald-type test will be estimated restricted to the null hypothesis (default and recommended) or unrestricted. Here we choose restricted (`"RML"`; unrestricted would be obtained with `"ML"`).
- The function h which is used to define the alternative as well as its inverse function h^{-1}. The function h is defined as `h = function(x){-log(x/(1-x))}` and its inverse function is defined as `h_inv = function(x){exp(-x)/(1+exp(-x))}`.

Then, the function `power_RET()` returns the sample size as part of an object of class *power.htest*.

```
ThreeArmedTrials::power_RET(experiment = 0.5, reference = 0.3, placebo = 0.2,
                            Delta = 0.8,
                            sig_level = 0.025,
                            power = 0.8,
                            allocation = w,
                            distribution = "binary",
                            h = function(x){-log(x/(1-x))},
                            h_inv = function(x){exp(-x)/(1+exp(-x))},
                            var_estimation = "RML")
```

```
##
##      Power calculation for Wald-type test (with restriced variance estimation) in three-arm
trial with binary endpoints
##
## Rate - Experimental = 0.5
##    Rate - Reference = 0.3
##      Rate - Placebo = 0.2
##                   n = 151
##           sig.level = 0.025
##               power = 0.8041344
```

```
##               Delta = 0.8
##          allocation = 0.4701987, 0.4105960, 0.1192053
##                nExp = 71
##                nRef = 62
##                nPla = 18
```

The object of class *power.htest* then contains the input information, the calculated sample size `n`, and the group specific sample sizes `nExp`, `nRef`, and `nPla`.

Analyzing a trial

Analyzing a three-arm trial with binary data will be demonstrated with data from a clinical trial in patients with depression which is included as the `remission` data set. The data set contains the three columns `experimental`, `reference`, and `placebo` with data from 86, 84, and 88 patients, respectively. In the experimental treatment group 43 patients (50%) went into remission, in the reference group 31 patients (36.9%) went into remission, and 26 patients (29.54%) in the placebo group went into remission.
We test the hypothesis specified with h as the logit function. The analysis will be performed with the function `test_RET()`. The distribution is specified by the argument `distribution = "binary"`. Moreover, we select the non-inferiority margin $\Delta = 0.8$ and the variance for the Wald-type test will be estimated restricted to the null hypothesis, i.e. `var_estimation = "RML"`.

```
library(ThreeArmedTrials)
noninf_test <- ThreeArmedTrials::test_RET(xExp = remission$experimental,
                                          xRef = remission$reference,
                                          xPla = remission$placebo,
                                          Delta = 0.8,
                                          var_estimation = "RML",
                                          distribution = "binary",
                                          h = function(x){-log(x/(1-x))},
                                          h_inv = function(x){exp(-x)/(1+exp(-x))})
```

The output of `test_RET()` is an object of class *h.test*.

```
print(noninf_test)
```

```
##
##  Wald-type test for the retention of effect hypothesis (with restriced
##  variance estimation) for the 'binary model'
##
## data:  remission$experimental, remission$reference, and remission$placebo
## T = -2.1183, p-value = 0.01707
## sample estimates:
##  Mean Exp  Mean Ref  Mean Pla
## 0.5000000 0.3690476 0.2954545
```

The output lists the test statistic, the p-value, and the group specific remission rates.

Planning and analyzing a study with the Poisson model

To demonstrate the functionality of the package *ThreeArmedTrials* for the Poisson model, we consider trials in epilepsy as an example. In trials in epilepsy a common endpoint is the number of seizures per person which is in general modelled as Poisson distributed.

Designing a trial

The following information is required to successfully calculate the sample size

- The vector of rates (μ_E, μ_R, μ_P) in the alternative. Here: $(\mu_E, \mu_R, \mu_P) = (15, 17, 20)$.
- A non-inferiority margin Δ. Here: $\Delta = 0.8$.
- A significance level and a power. Here: $\alpha = 0.025$ and $\text{Power} = 0.8$.
- A sample size allocation. Here: $n_E : n_R : n_P = 1 : 1 : 1$, i.e. a balanced design.
- Whether the variance in the Wald-type test will be estimated restricted to the null hypothesis (default and recommended) or unrestricted. Here we choose restricted (`"RML"`; unrestricted would be obtained with `"ML"`).

To calculate the sample size, we use the `power_RET()` function with the parameter `distribution = "poisson"` to specify that our data is modeled as Poisson distributed. The output of the function `power_RET()` is an object of class *power.htest*:

```
library(ThreeArmedTrials)
power_RET(experiment = 15, reference = 17, placebo = 20,
          Delta = 0.8, sig_level = 0.025, power = 0.8,
          allocation = c(1, 1, 1) / 3,
          var_estimation = "RML",
          distribution = "poisson")
```

```
##
##      Power calculation for Wald-type test (with restriced variance estimation) in three-arm
trial with Poisson endpoints
##
## Rate - Experiment = 15
##  Rate - Reference = 17
##    Rate - Placebo = 20
##                 n = 96
##         sig.level = 0.025
##             power = 0.8047648
##             Delta = 0.8
##        allocation = 0.3333333, 0.3333333, 0.3333333
##              nExp = 32
##              nRef = 32
##              nPla = 32
```

First the setup is described with a note. Then, the output lists of the numeric input parameters as well as the calculated total sample size `n` and the group sample sizes `nExp`, `nRef`, and `nPla`.

Here, we planned the sample size for a balanced design, i.e. $n_E : n_R : n_P = 1 : 1 : 1$. Alternatively, the sample size allocation could be calculated for a given alternative such that the power is maximized. The resulting sample size allocation is referred to as the optimal sample size allocation. The optimal sample size allocation can be calculated with the function `opt_alloc_RET()`. Thereto, the vector of rates (μ_E, μ_R, μ_P), the non-inferiority margin Δ, and the distribution must be specified. For the example from above, the optimal sample size allocation can be calculated by:

```
opt_alloc_RET(experiment = 15,
              reference = 17,
              placebo = 20,
              Delta = 0.8,
              distribution = "poisson")
```

```
## [1] 0.4801678 0.4089422 0.1108900
```

The function `opt_alloc_RET()` returns the vector with the optimal sample size allocation. The first element refers to the experimental treatment, the second to the reference, and the third to the placebo.

Analyzing a trial

As an example for a trial with Poisson data we considered the dataset `seizures` which is included in the *ThreeArmedTrials* package. The `seizures` dataset contains the number of seizures per patient for three different add-on treatments, i.e. placebo, reference, and experimental treatment, evaluating an anti-epileptic drug. The dataset includes 32 patients.

First six entries of seizures dataset

	pla	ref	exp
1	21	13	14
2	12	14	13
3	21	5	11
4	26	19	18
5	20	17	14
6	17	21	18

For the `seizures` data set non-inferiority of the experimental treatment compared to the reference will be tested with a non-inferiority margin of $\Delta = 0.8$. The standard deviation for the Wald-type test will be estimated using a maximum-likelihood estimator restricted to the parameter space of the null hypothesis. Thus, we set the parameter `var_estimation = "RML"`. For the unrestricted estimator, `var_estimation = "ML"` can be used. The hypothesis test is performed with the function `test_RET()`:

```
noninf_test <- ThreeArmedTrials::test_RET(xExp = seizures$exp,
                                          xRef = seizures$ref,
                                          xPla = seizures$pla,
                                          Delta = 0.8,
                                          var_estimation = "RML",
                                          distribution = "poisson")
```

The output of the function `test_RET()` is an object of class *h.test*.

```
print(noninf_test)
```

```
##
##  Wald-type test for the retention of effect hypothesis (with restriced
##  variance estimation) for the 'poisson model'
##
## data:  seizures$exp, seizures$ref, and seizures$pla
## T = -2.1824, p-value = 0.01454
## sample estimates:
## Mean Exp Mean Ref Mean Pla
## 15.15625 16.31250 20.56250
```

The output starts with a note describing the setup. Then, the names of the data variables are listed followed by the test statistic `T` and the p-value. The output is concluded with the group specific means.

Analyzing a study with the non-parametric model

In this section the focus is on analyzing a three-arm study in the *gold standard* design using the non-parametric method, i.e. a studentized permutation test. We focus on the exemplary data set `GElesions`.

First six entries of GElesions dataset

	experimental	reference	placebo
1	0	0	3
2	10	6	0
3	0	3	15
4	2	0	20
5	0	8	0
6	0	0	0

We test non-inferiority of the experimental treatment compared to the reference with a margin of $\Delta = 0.8$. Additionally, we must specify the number of permutations considered when calculating the rejection area. Here, we choose `n_perm = 50000`. As before, hypothesis is test with the function `test_RET()`:

```
set.seed(12345)
noninf_test <- ThreeArmedTrials::test_RET(xExp = GElesions$experimental,
                                          xRef = GElesions$reference,
                                          xPla = GElesions$placebo,
                                          Delta = 0.8,
                                          n_perm = 50000,
                                          distribution = "nonparametric")
```

The output of the function `test_RET()` is an object of class *h.test*.

```
print(noninf_test)
```

```
##
##  Studentized permutation test for the retention of effect hypothesis
```

```
##
## data:  GElesions$experimental, GElesions$reference, and GElesions$placebo
## T = -0.45875, p-value = 0.345
## sample estimates:
## Mean Exp Mean Ref Mean Pla
##     2.82     2.78     5.56
```

Here, we see that due to the p-value of about 34%, non-inferiority could not be shown with a commonly considered significance level.

References

- I. Pigeot, J. Schäfer, J. Röhmel, D. Hauschke. *Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo.* Statistics in Medicine, 22:883-899 (2003).
- M. Kieser, T. Friede. *Planning and analysis of three-arm non-inferiority trials with binary endpoints.* Statistics in Medicine, 26:253-273 (2007).
- M. Hasler, R. Vonk, L.A. Hothorn. *Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity.* Statistics in Medicine, 27:490-503 (2008).
- M. Mielke, A. Munk, and A. Schacht. *The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints.* Statistics in Medicine 27:5093-5110 (2008).
- M. Mielke and A. Munk. *The assessment and planning of non-inferiority trials for retention of effect hypotheses-towards a general approach.* arXiv preprint arXiv:0912.4169 (2009).
- M. Mielke. *Maximum Likelihood Theory for Retention of Effect Non-inferiority Trials*
- T. Mütze, A. Munk, T. Friede. *Design and analysis of three-arm trials with negative binomially distributed endpoints*. Statistics in Medicine. 35:505-521 (2016).
- T. Mütze, F. Konietschke, A. Munk, T. Friede. *A studentized permutation test for three-arm trials in the `gold standard' design*. Statistics in Medicine. 36:883-898 (2017).