# Package 'PRIMEplus'

January 20, 2025

**Title** Study Design for Immunotherapy Clinical Trials

**Version** 1.0.16

**Date** 2024-01-09

**Description** Perform sample size, power calculation and subsequent analysis for Immuno-oncology (IO) trials composed of responders and non-responders.

**Maintainer** Bill Wheeler <wheelerb@imsweb.com>

**Depends** R (>= 3.5), survival, msm

**License** GPL-2

**NeedsCompilation** yes

**Author** Zhenzhen Xu [aut],
Yongsoek Park [aut],
Zhu Bin [aut],
Bill Wheeler [cre]

**Repository** CRAN

**Date/Publication** 2024-01-10 14:03:13 UTC

## Contents

## PRIMEplus-package *Study design for immunotherapy clinical trials*

### Description

Perform sample size, power calculation and subsequent analysis for Immuno-oncology (IO) trials composed of responders and nonresponders.

### Details

This package is an extension of the Immunotherapy.Design R package but allows for response categories of more than two categories among the treatment population, such as complete responders, partial responders, as well as non-responders.

### Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

### References

Xu, Z., Park, Y., Liu, K. and Zhu, B. Treating non-responders: pitfalls and implications for cancer immunotherapy trial design. Journal of Hematology & Oncology 13, 20 (2020).

Xu, Z., Zhu, B. and Park, Y. Designing immuno-oncology clinical trials composed of responders and nonresponders. Statistics in Medicine. (Under Revision).

## data *Data for examples*

### Description

Data for examples.

### Details

A data frame used in the examples.

### Examples

```
data(data, package="PRIMEplus")

# Display some of the data
data[1:5, ]
```

`generate_data` *Simulated data*

## Description

Generate simulated data

## Usage

```
generate_data(N=500, rand_ratio=0.5, effect_p=0.6, enroll_rate=380*0.25/6,
              lambda1=0.117, HR=0.5, tau=12*5, t1=1)
```

## Arguments

| | |
|---|---|
| `N` | Maximum sample size |
| `rand_ratio` | Allocation ratio |
| `effect_p` | Vector for proportion of responders in the treatment arm at baseline (see details) |
| `enroll_rate` | Enrollment rate in subjects per month |
| `lambda1` | Baseline hazard in terms of months |
| `HR` | Vector of hazard ratios for responders against controls (see details) |
| `tau` | Total study duration |
| `t1` | Delayed duration |

## Details

The length and order of `effect_p` must be the same as `HR`. Both of these vectors should contain information only for the groups of responders. For example, if there are full responders and partial responders, then `effect_p` and `HR` would be vectors of length two.

## Value

A data frame with columns:

| Name | Description |
|---|---|
| id | id variable |
| trt | treatment allocation: 1 = treatment arm |
| D1 | patient's response status for group 1 |
| D2 | patient's response status for group 2 |
| Dm | patient's response status for non-responders |
| tau | total study duration |
| enroll_time | patients' enrollment times |
| time_to_event | patients' event times |
| event_status | censoring indicator |
| X | observational time |
| t1 | delayed duration |

### Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

### Examples

```
data <- generate_data()
data[1:5, ]
```

---

## getHazard *Compute initial estimates for the baseline hazard*

---

### Description

Calls the coxph function to compute initial estimates for the baseline hazard

### Usage

```
getHazard(time, treatment, event_status, t.fail.o=NULL)
```

### Arguments

| | |
|---|---|
| time | Vector of times. |
| treatment | Binary vector of treatments (1=subject received treatment). |
| event_status | Binary vector of event status (1=subject experienced an event). |
| t.fail.o | NULL or vector of event times. |

### Value

List containing the baseline hazards ordered by the event times.

### Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov> and Bin Zhu <bin.zhu@nih.gov>

### See Also

PRIMEplus.EM

### Examples

```
data(data, package="PRIMEplus")
lambda0 <- getHazard(data[, "X"], data[, "trt"], data[, "event_status"])$hazard
lambda0[1:10]
```

PRIMEplus.EM *EM algorithm*

## Description

EM algorithm

## Usage

```
PRIMEplus.EM(data, effect_p, beta0, time.var="X", trt.var="trt",
       status.var="event_status", id.var="id", t1=1, lambda0=NULL,
       stopTol=1e-4, maxiter=100000, print=0)
```

## Arguments

| | |
|---|---|
| data | Data frame or matrix containing a time-to-event variable (time.var), a treatment variable (trt.var), and a censoring variable (status.var). |
| effect_p | Vector of proportions for groups of responders in the treatment arm at baseline (see details). |
| beta0 | Vector or matrix of initial estimates for the log-hazard ratios (see details). |
| time.var | Time-to-event variable name in data. The default is "X". |
| trt.var | Binary treatment variable name in data coded as 0 for controls and 1 for subjects that received treatment. |
| status.var | Name of the binary censoring variable in data coded as 0 for censored subjects and 1 for subjects that experienced an event. |
| id.var | NULL or subject id variable in data. The default is "id". |
| t1 | Delayed duration. The default is 1. |
| lambda0 | NULL or vector of initial estimates for the baseline hazards corrsponding to the ordered event times. |
| stopTol | Stopping tolerance. The default is 1e-4. |
| maxiter | Maximum number of iterations. The default is 100000. |
| print | 0-2 to print information. Larger values will print more information. The default is 0. |

## Details

The EM algorithm is sensitive to the initial values of the log-hazard ratios (beta0), so different initial estimates should be tried to ensure the maximum log-likelihood is obtained. Thus, beta0 can be a vector or matrix, where in the case of a matrix, each row corresponds to a different set of intial estimates. Each set of initial estimates must contain distinct non-zero values. The length and order of effect_p must be the same as the columns of beta0. Both of these should contain information only for the groups of responders. For example, if there are full responders and partial responders, then effect_p would be a vector of length two, and beta0 would be a vector of length two or a matrix with two columns.

## Value

A list containing the objects:

| Name | Description |
|---|---|
| converged | TRUE if EM algorithm converged |
| beta | final log(hazard ratio) estimates of responders versus controls |
| lambda | final estimates of baseline hazards |
| probResponder | estimated probability of a subject being a responder |
| loglike | log-likelihood value at the final estimates |
| loglike.marg | marginal log-likelihood value at the final estimates |

## Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

## See Also

getHazard, PRIMEplus.LRT

## Examples

```
data(data, package="PRIMEplus")
effp  <- c(0.3, 0.3)
beta0 <- c(log(0.4), log(0.6))
ret   <- PRIMEplus.EM(data, effp, beta0)
ret$beta
```

---

# PRIMEplus.LRT *Likelihood Ratio Test*

---

## Description

PRIMEplus likelihood ratio test

## Usage

```
PRIMEplus.LRT(data, effect_p, beta0, time.var="X", trt.var="trt",
      status.var="event_status", id.var="id", t1=1, lambda0=NULL,
      stopTol=1e-4, maxiter=100000, print=0)
```

## Arguments

| | |
|---|---|
| data | Data frame or matrix containing a time-to-event variable (time.var), a treatment variable (trt.var), and a censoring variable (status.var). |
| effect_p | Vector of proportions for groups of responders in the treatment arm at baseline (see details). |

| | |
|---|---|
| beta0 | Vector or matrix of initial estimates for the log-hazard ratios (see details). |
| time.var | Time-to-event variable name in data. The default is "X". |
| trt.var | Binary treatment variable name in data coded as 0 for controls and 1 for subjects that received treatment. |
| status.var | Name of the binary censoring variable in data coded as 0 for censored subjects and 1 for subjects that experienced an event. |
| id.var | NULL or subject id variable in data. The default is "id". |
| t1 | Delayed duration. The default is 1. |
| lambda0 | NULL or vector of initial estimates for the baseline hazards corrsponding to the ordered event times. |
| stopTol | Stopping tolerance. The default is 1e-4. |
| maxiter | Maximum number of iterations. The default is 100000. |
| print | 0-2 to print information. Larger values will print more information. The default is 0. |

## Details

The EM algorithm is sensitive to the initial values of the log-hazard ratios (beta0), so different initial estimates should be tried to ensure the maximum log-likelihood is obtained. Thus, beta0 can be a vector or matrix, where in the case of a matrix, each row corresponds to a different set of intial estimates. Each set of initial estimates must contain distinct non-zero values. The length and order of effect_p must be the same as the columns of beta0. Both of these should contain information only for the groups of responders. For example, if there are full responders and partial responders, then effect_p would be a vector of length two, and beta0 would be a vector of length two or a matrix with two columns.

## Value

A list containing the objects:

| Name | Description |
|---|---|
| converged | TRUE if EM algorithm converged |
| beta | final log(hazard ratio) estimates of responders versus controls |
| lambda | final estimates of baseline hazards |
| probResponder | estimated probability of a subject being a responder |
| loglike | log-likelihood value at the final estimates |
| loglike.marg | marginal log-likelihood value at the final estimates |
| loglike.marg.0 | marginal log-likelihood value under the null hypothesis |
| LRT | likelihood-ratio test statistic |
| pvalue | p-value of the likelihood ratio test |

## Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

## See Also

PRIMEplus.EM

## Examples

```
data(data, package="PRIMEplus")
effp  <- c(0.3, 0.3)
beta0 <- c(log(0.4), log(0.6))
ret   <- PRIMEplus.LRT(data, effp, beta0)
ret$LRT
ret$pvalue
```

---

PRIMEplus.Power *Power*

---

## Description

Compute the power using LRT Re-randomization test

## Usage

```
PRIMEplus.Power(nmax=500, rand_ratio=0.5, effect_p=0.6,
    enroll_rate=380*0.25/6, lambda1=0.117, HR=0.5, tau=12*5, t1=1,
    maxiter=100000, stopTol=1e-4, alpha=0.05, num_rand=1000, nsim=10000,
    print=0, min.sample.size=50, min.n.event=5, min.per.trt=0.25)
```

## Arguments

| | |
|---|---|
| nmax | Sample size |
| rand_ratio | Probability of assignment to treatment arm |
| effect_p | Vector for proportion of responders in the treatment arm at baseline (see details) |
| enroll_rate | Enrollment rate in subjects per month |
| lambda1 | Baseline hazard in terms of months |
| HR | Vector of hazard ratios for responders against controls (see details) |
| tau | Total study duration |
| t1 | Delayed duration in months |
| maxiter | Maximum number of iterations in the EM algorithm. The default is 100000. |
| stopTol | Stopping tolerance in the EM algorithm. The default is 1e-4. |
| alpha | Significance level. The default is 0.05. |
| num_rand | The number of replications in the re-randomization test. The default is 1000. |
| nsim | The number of simulations. The default is 1000. |
| print | 0 or 1 to print information. The default is 0. |
| min.sample.size | Minimum sample size. The default is 50. |
| min.n.event | Minimum number of events. The default is 5. |
| min.per.trt | Minimum proportion of controls and treated subjects. The default is 0.25. |

## Details

The length and order of `effect_p` must be the same as `HR`. Both of these vectors should contain information only for the groups of responders. For example, if there are full responders and partial responders, then `effect_p` and `HR` would be vectors of length two.

For each simulation, a simulated data set is created from the `generate_data` function and then an estimated p-value is computed by calling `PRIMEplus.ReRand.LRT`. The power is calculated as the proportion of iterations whose estimated p-value was less than or equal to `alpha`.

## Value

A list containing the power and the number of simulated datasets used in the calculation.

## Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

## See Also

`PRIMEplus.ReRand.LRT`

---

# `PRIMEplus.ReRand.LRT` *Randomization Test*

---

## Description

PRIMEplus randomization likelihood ratio test

## Usage

```
PRIMEplus.ReRand.LRT(data, effect_p, beta0, time.var="X", trt.var="trt",
                     status.var="event_status", id.var="id", t1=1, lambda0=NULL,
                          stopTol=1e-4, maxiter=100000, print=0, num_rand=1000)
```

## Arguments

| | |
|---|---|
| `data` | Data frame or matrix containing a time-to-event variable (`time.var`), a treatment variable (`trt.var`), and a censoring variable (`status.var`). |
| `effect_p` | Vector of proportions for groups of responders in the treatment arm at baseline (see details). |
| `beta0` | Vector or matrix of initial estimates for the log-hazard ratios (see details). |
| `time.var` | Time-to-event variable name in `data`. The default is "X". |
| `trt.var` | Binary treatment variable name in `data` coded as 0 for controls and 1 for subjects that received treatment. |
| `status.var` | Name of the binary censoring variable in `data` coded as 0 for censored subjects and 1 for subjects that experienced an event. |

| | |
|---|---|
| id.var | NULL or subject id variable in data. The default is "id". |
| t1 | Delayed duration. The default is 1. |
| lambda0 | NULL or vector of initial estimates for the baseline hazards corrsponding to the ordered event times. |
| stopTol | Stopping tolerance. The default is 1e-4. |
| maxiter | Maximum number of iterations. The default is 100000. |
| print | 0-2 to print information. Larger values will print more information. The default is 0. |
| num_rand | The number of randomizations. The default is 1000. |

## Details

The EM algorithm is sensitive to the initial values of the log-hazard ratios (beta0), so different initial estimates should be tried to ensure the maximum log-likelihood is obtained. Thus, beta0 can be a vector or matrix, where in the case of a matrix, each row corresponds to a different set of intial estimates. Each set of initial estimates must contain distinct non-zero values. The length and order of effect_p must be the same as the columns of beta0. Both of these should contain information only for the groups of responders. For example, if there are full responders and partial responders, then effect_p would be a vector of length two, and beta0 would be a vector of length two or a matrix with two columns. The initial values are only used for the observed data; each randomization uses the MLE estimates as initial estimates.

## Value

A list containing the objects:

| Name | Description |
|---|---|
| pvalue.LRT | p-value of the randomization test based on the likelihood ratio test |
| pvalue.loglike.marg | p-value of the randomization test based on the marginal likelihood |
| n.randomizations | the number of randomizations that the p-values are based on |

## Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

## See Also

PRIMEplus.LRT

## Examples

```
data(data, package="PRIMEplus")
effp  <- c(0.3, 0.3)
beta0 <- c(log(0.4), log(0.6))
ret   <- PRIMEplus.ReRand.LRT(data, effp, beta0, num_rand=100)
ret
```

PRIMEplus.SampleSize *Sample Size*

## Description

Compute the sample size for a given power

## Usage

```
PRIMEplus.SampleSize(power=0.8, rand_ratio=0.5, effect_p=0.6, enroll_rate=380*0.25/6,
            lambda1=0.117, HR=0.5, tau=12*5, t1=1, maxiter=100000, stopTol=1e-4,
              alpha=0.05, num_rand=1000, nsim=10000, min.N=100, max.N=700,
              tol.power=0.01, tol.N=1, print=1,
              min.sample.size=50, min.n.event=5, min.per.trt=0.25)
```

## Arguments

| | |
|---|---|
| power | The desired power. The default is 0.8. |
| rand_ratio | Allocation ratio |
| effect_p | Vector for proportion of responders in the treatment arm at baseline (see details) |
| enroll_rate | Enrollment rate in subjects per month |
| lambda1 | Baseline hazard in terms of months |
| HR | Vector of hazard ratios for responders against controls (see details) |
| tau | Total study duration |
| t1 | Delayed duration |
| maxiter | Maximum number of iterations in the EM algorithm. The default is 100000. |
| stopTol | Stopping tolerance in the EM algorithm. The default is 1e-4. |
| alpha | Significance level. The default is 0.05. |
| num_rand | Number of replications in the re-randomization test. The default is 1000. |
| nsim | Number of simulations in computing power (**see Details**). The default is 10000. |
| min.N | Lower bound for the sample size. The default is 100. |
| max.N | Upper bound for the sample size. The default is 700. |
| tol.power | Stopping tolerance for the power. The default is 0.01. |
| tol.N | Stopping tolerance for the sample size. The default is 1. |
| print | 0 or 1 to print information. The default is 1. |
| min.sample.size | Minimum sample size. The default is 50. |
| min.n.event | Minimum number of events. The default is 5. |
| min.per.trt | Minimum proportion of controls and treated subjects. The default is 0.25. |

## Details

The length and order of `effect_p` must be the same as `HR`. Both of these vectors should contain information only for the groups of responders. For example, if there are full responders and partial responders, then `effect_p` and `HR` would be vectors of length two.

This uses a bisection method to estimate the sample size. At each iteration, the estimated power `power_est` is computed using `PRIMEplus.Power` for a given sample size holding all other parameters fixed. The algorithm terminates when `abs(power - power_est) <= tol.power` or when the length of the estimated interval containing the sample size is less than or equal to `tol.N`.

**NOTE:**
**It is important to note that the power for a given sample size is estimated by running a simulation. Thus, by setting a different seed, a different result may be returned. Therefore, to ensure a more precise estimated sample size, set the option `nsim` to a large value and/or run this function several times by setting different seeds and examine the distribution of returned sample sizes.**

## Value

A list containing the sample size and power.

## Author(s)

Zhenzhen Xu <Zhenzhen.Xu@fda.hhs.gov>, Yongsoek Park <yongpark@pitt.edu> and Bin Zhu <bin.zhu@nih.gov>

## See Also

`PRIMEplus.Power`

# Index