

# Guideline for codes

## Restricted mean survival time to estimate an intervention effect in a cluster randomized trial

Floriane Le Vilain–Abraham, Elsa Tavernier, Etienne Dantan,  
Solène Desmée, Agnès Caille

### Contents

<b>0</b>	<b>Introduction</b>	<b>1</b>
<b>1</b>	<b>Generation of the datasets and statistical analysis</b>	<b>1</b>
<b>2</b>	<b>Estimation of the performance measures</b>	<b>5</b>

## 0 Introduction

The code is divided into two folders: one for the scenarios under proportional-hazards (PH) assumption and one under non-proportional-hazards (NPH) assumption. For both cases, there are 2 steps:

1. Generation of the datasets and statistical analysis
2. Estimations of the performance measures

The following sections describe how to use the R code for each step.

## 1 Generation of the datasets and statistical analysis

The `main_DA_PH.R` (for the PH assumption) and `main_DA_NPH.R` (for the NPH assumption) files are R codes to generate 1000 simulated datasets of a predefined scenario and analyse each dataset with the 5 methods ( $KM_{indep}$ ,  $PV_{indep}$ ,  $KM_{clust}$ ,  $PV_{ECM}$  and  $PV_{ICM}$ ). For the permutation test, see the subsection [1](#).

The necessary packages are: `doParallel` (line 14) and `doRNG` (line 15). The necessary R functions, saved in a R file with the same name, are loaded in the main script (`main_DA_PH.R` or `main_DA_NPH.R`):

- `sim_PH` or `sim_NPH` (line 20): simulate a dataset and estimate the difference in RMST, its variance and the 95% confidence interval with the 5 methods
- `generate_data_PH` or `generate_data_NPH` (line 21): generate one dataset
- `generate_cluster_PH` or `generate_cluster_NPH` (line 22): generate time-to-event data for one cluster
- `RMST_pseudo` (line 23): analysis of one dataset with the pseudo-value regression-based methods (accounting and not accounting for clustering)
- `RMST_KM` (line 24): analysis of one dataset with the Kaplan-Meier-based methods (accounting and not accounting for clustering)
- `variance_bootstrap` (line 25): resample one dataset and estimate the difference in RMST with the Kaplan-Meier-based method (function to estimate the variance of the difference in RMST using bootstrap method)

The `RMST_pseudo`, `RMST_KM` and `variance_bootstrap` functions are identical for PH and NPH assumptions.

Before running, you have to specify in the main file (`main_DA_PH.R` or `main_DA_NPH.R`):

- The directory where you saved all the R files with the necessary R functions detailed previously (line 19).
- The parameters of the scenario (line 30 in the dataframe `table_parameter`).
- The horizon time (line 41). Here, it is set at 365 days. All time are expressed in days.
- The directory where the 1000 simulated datasets and the estimations of the difference in RMST, variances and 95% confidence intervals will be saved (line 47). The 1000 datasets will be saved in a same automatically created folder. The estimations of the difference in RMST, variance and 95% confidence interval for the 5 methods and for the 1000 simulated datasets will be saved in one txt file. Table 1 and 2 described the content of one simulated dataset and the txt file with all the estimations, respectively.
- The number of cores for the parallelisation (line 76).

Colnames	Variables
time	$X_{lk}$ the observed time of the individual $l$ from the cluster $k$
arm	arm of the clusters $k$ (= 1 for the intervention group, 0 for control group)
cluster	cluster's identifiant $k$
id_patient	patient's identifiant $l$
status	$\delta_{lk}$ the event indicator of the individual $l$ from the cluster $k$ (= 1 for death, 0 for censor)

Table 1: Content of one simulated dataset

Colnames	Variables
<i>Parameters used to simulate the dataset</i>	
d	dataset number
K	total number of clusters
m	mean cluster size
HR	true HR
t_delay	change point ( <u>only for NPH assumption</u> )
tau	Kendall's tau
censoring	censoring rate
<i>Analysis</i>	
clustering	if the method accounts (= "yes") or does not account (= "no") for clustering
delta.rmst	estimation of the difference in RSMT
var	variance estimation
ci.low	lower bound of the 95% confidence interval
ci.up	upper bound of the 95% confidence interval
method	type of method used for the analysis
t_star	horizon time of the analysis

Table 2: Content of txt file with the estimations of the difference in RMST, variance and 95% confidence intervals for the 5 methods and for the 1000 simulated datasets

### ***Permutation test for the pseudo-values regression-based method: Statistical analysis***

The permutation tests for the pseudo-values regression-based method with a independent and exchangeable working correlation matrix are performed separately. All the needed functions are available in the *Permutation test* folder (*/Proportional-hazards/1 - Genration of the datasets and statistical analysis/*).

The `main_permutation.R` file is the R code to analyse each simulated datasets with permutation test. The necessary packages are: `doParallel` (line 14) and `doRNG` (line 15).

The necessary R functions, saved in a R file with the same name, are loaded in the main file (`main_permutation.R`):

- `sim_permutation` (line 20): to estimate the p-values for the pseudo-values regression-based methods
- `permutation` (line 21): to analyse of one dataset with the permutation test
- `permutation_test` (line 22): to estimate the test statistic for one permuted dataset

Before running, you have to specify in the `main_permutation.R` file:

- The directory where you saved all the R files with the necessary R functions detailed previously (line 19).
- The parameters of the scenarios (line 27 in the dataframe `table_parameter`).
- The horizon time (line 38). Here, it is set at 365 days. All time are expressed in days.
- The directory where the estimations of the p-values will be saved (line 45). The estimations of the p-values for the 1000 simulated datasets will be saved in one txt file. Table 3 describes this txt file.
- The number of cores for the parallelisation (line 76).

You also have to specify in the `sim_permutation.R` file, the directory where the simulated datasets have been saved (line 24).

Colnames	Variables
<i>Parameters used to simulate the dataset</i>	
<code>d</code>	dataset number
<code>K</code>	total number of clusters
<code>m</code>	mean cluster size
<code>HR</code>	true HR
<code>tau</code>	Kendall's tau
<code>censoring</code>	censoring rate
<i>Analysis</i>	
<code>matrix</code>	type of working correlation matrix used for the analysis ("ind" = independent, "exch" = exchangeable)
<code>var</code>	estimated p-value
<code>t_star</code>	horizon time of the analysis

Table 3: Content of txt file with the estimations of p-value for pseudo-values regression for the 1000 simulated datasets

## 2 Estimation of the performance measures

The `main_PM_PH.R` (for the PH assumption) and `main_PM_NPH.R` (for the NPH assumption) files are R codes to estimate the performance measures for the 5 methods. ( $KM_{indep}$ ,  $PV_{indep}$ ,  $KM_{clust}$ ,  $PV_{ECM}$  and  $PV_{ICM}$ ). For the permutation test, the code see the subsection [2](#).

No package is necessary. The necessary R functions, saved in a R file with the same name, are loaded in the main script:

- `pm_estimation_PH` or `pm_estimation_NPH` (line 15): estimation of the performance measures for the 5 methods
- `performance_measures` (line 16): estimation of the performance measure for one methode
- `true_rmst_difference_PH` or `true_rmst_difference_NPH` (line 17): computation of the true difference in RSMT
- `survival_function_PH` or `survival_function_NPH` (line 18): true survival function

Before running, you have to specify:

- The directory where you saved all the R files with the necessary R functions detailed previously (line 14).
- The directory and the name of the txt file where the estimations of the difference in RMST, variance and 95% confidence interval for the 5 methods and for the 1000 datasets have been saved in the step 1 (line 23).

When you run the main R script, the performance measures are summarized in a `data.frame`. The content of this `data.frame` is detailed in Table [4](#).

Colnames	Variables
method	method
tstar	horizon time for the analysis
K	total number of clusters
m	mean cluster size
HR	true HR
tau	Kendall's tau
censoring	censoring rate
RB	relative bias
RE	relative error
coverage	coverage rate
rejection.rate	type I error rate ( <u>for the null hypothesis</u> , i.e. absence of intervention effect)
D	number of simulation iterations that converge

Table 4: Content of the data.frame obtained when running the main scripts `main_PM_PH.R` (for the PH assumption) or `main_PM_NPH.R` (for the NPH assumption)

### ***Permutation test for the pseudo-values regression-based method: Estimation of the type I error rate***

The permutation test for the pseudo-values regression is performed separately. All the needed functions are available in the *Permutation test* folder (*/Proportional-hazards/2 - Estimation of the performance measures/*).

The `main_PM_permutation.R` file is R codes to estimate the type I error rate for the pseudo-values regression-based method. No package is necessary. The following necessary R function, saved in a R file with the same name, are loaded in the main script:

- `pm.estimation_permuation` (line 15): estimation of the type I error rate with the permutation test for the pseudo-values regression based-methods

Before running, you have to specify in the `main_PM_permutation.R` file:

- The directory where you saved all the R files with the necessary R functions detailed previously (line 14).
- The directory and the name of the txt file where the estimations of the p-values for the the permutation test for the pseudo-values regression-based methods for the 1000 datasets have been saved in the step 1 (line 21).

When you run the main R script, the type I error rates are summarized in a data.frame. The content of this data.frame is detailed in [Table 5](#)

Colnames	Variables
<code>tstar</code>	horizon time for the analysis
<code>K</code>	total number of clusters
<code>m</code>	mean cluster size
<code>HR</code>	true HR
<code>tau</code>	Kendall's tau
<code>censoring</code>	censoring rate
<code>matrix</code>	type of working correlation matrix used for the analysis
<code>tie</code>	type I error rate

Table 5: Content of the data.frame obtained when running the main scripts `main_PM_PH.R` (for the PH assumption) or `main_PM_NPH.R` (for the NPH assumption)