

# Package: MSCA (via r-universe)

February 20, 2025

**Title** Clustering of Multiple Censored Time-to-Event Endpoints

**Version** 1.0

**Description** Provides basic tools for computing clusters of instances described by multiple time-to-event censored endpoints. From long-format datasets, where one instance is described by one or more records of events, a procedure is used to compute state matrices. Then, from state matrices, a procedure provides optimised computation of the Jaccard distance between instances. The library is currently in development, and more options and tools allowing graphical representation of typologies are expected. For methodological details, see our methodological paper: Delord M, Douiri A (2025) [doi:10.1186/s12874-025-02476-7](https://doi.org/10.1186/s12874-025-02476-7).

**License** GPL-3

**Imports** dplyr, Matrix, Rcpp, RcppParallel, rlang, stats, tibble, utils

**LinkingTo** Rcpp, RcppArmadillo, RcppParallel

**Encoding** UTF-8

**RoxygenNote** 7.3.2

**NeedsCompilation** yes

**Author** Marc Delord [aut, cre]  
(<https://orcid.org/0000-0002-0455-6749>)

**Maintainer** Marc Delord <mdelord@gmail.com>

**Depends** R (>= 3.5.0)

**Date/Publication** 2025-02-19 14:30:13 UTC

**Config/pak/sysreqs** make

**Repository** <https://marc-75.r-universe.dev>

**RemoteUrl** <https://github.com/cran/MSCA>

**RemoteRef** HEAD

**RemoteSha** 649da687881af065ed84a5cc4d24cd3afaa760f9

## Contents

EHR . . . . .	2
make_state_matrix . . . . .	2
MSCA . . . . .	3
msca . . . . .	4
<b>Index</b>	<b>5</b>

---

EHR	<i>Description of the EHR dataset</i>
-----	---------------------------------------

---

### Description

This is a toy dataset to illustrate the use of the the MSCA library

### Format

A data frame with 3000 records and 3 variables:

- link\_id** Unique patient identifier
- reg** Registered long-term condition
- aos** Age at onset of the registered long-term condition

### Source

Toy dataset

### Examples

```
# Load the dataset
data(EHR)

# Display the first few rows
head(EHR)
```

---

make_state_matrix	<i>Create a State Matrix</i>
-------------------	------------------------------

---

### Description

This function generates state matrices based on the given dataset, with options for defining ID columns, LTC columns, and matrix dimensions. It splits the data, processes it by group, and computes matrices for each subset of data.

**Usage**

```
make_state_matrix(  
  data,  
  id = "link_id",  
  ltc = "reg",  
  aos = "aos",  
  k = 100,  
  l = 111,  
  fail_code = "death",  
  cens_code = "cens"  
)
```

**Arguments**

data	A data frame or tibble containing the dataset long format
id	A string representing the column name of patients ID
ltc	A string representing the column name of the event (long term conditions)
aos	A string representing the column name of age at onset of considered long term conditions
k	An integer representing the group size when splitting the dataset (usefull for large datasets)
l	An integer representing the number of rows of the state matrices. It usualle represent the maximum age of patients
fail_code	A string representing the code used for failure (death)
cens_code	A string representing the code used for censoring

**Details**

The function processes the data by splitting it based on the `id` and `ltc` columns. It computes matrices for each subset, with rows representing different states and columns representing unique IDs.

**Value**

A list of state matrices, one per patient

**Description**

The MSCA library provides basic tools for computing clusters of instances described by multiple time-to-event censored endpoints. From long format datasets, where one instance is described by one or more record of events, a procedure is used to compute states matrices. Then, from states matrices, a procedure provides optimised computation of the Jacard distance between instances. The library is currently in development and more options and tools allowing graphical representation of typologies are expected.

**Details**

This package is under active development

**Author(s)**

Marc Delord

**Reference:** Delord M, Douiri A (2025) "Multiple states clustering analysis (MSCA), an unsupervised approach to multiple time-to-event electronic health records applied to multimorbidity associated with myocardial infarction." *BMC Med Res Methodol.* 25(1):32. <doi:10.1186/s12874-025-02476-7>

Maintainer: Marc Delord <mdelord@gmail.com>

---

msca

*Compute Jaccard Index Using Parallel Rcpp*

---

**Description**

This function takes a matrix, removes rows where all values are zero (or NA), and computes the Jaccard index using an Rcpp parallel function.

**Usage**

```
msca(smats)
```

**Arguments**

smats            A numeric matrix where Jaccard Index will be computed between the columns.

**Value**

A matrix containing the Jaccard index between columns.

# Index

\* **clustering**

MSCA, [3](#)

EHR, [2](#)

make\_state\_matrix, [2](#)

MSCA, [3](#)

msca, [4](#)

MSCA-package (MSCA), [3](#)