

# Package ‘crrSC’

October 12, 2022

**Title** Competing Risks Regression for Stratified and Clustered Data

**Version** 1.1.2

**Author** Bingqing Zhou and Aurelien Latouche

**Description** Extension of 'cmprsk' to Stratified and Clustered data.

A goodness of fit test for Fine-Gray model is also provided.

Methods are detailed in the following articles: Zhou et al. (2011) <[doi:10.1111/j.1541-0420.2010.01493.x](https://doi.org/10.1111/j.1541-0420.2010.01493.x)>,

Zhou et al. (2012) <[doi:10.1093/biostatistics/kxr032](https://doi.org/10.1093/biostatistics/kxr032)>,

Zhou et al. (2013) <[doi:10.1002/sim.5815](https://doi.org/10.1002/sim.5815)>.

**Depends** survival

**Maintainer** Aurelien Latouche <[aurelien.latouche@cnam.fr](mailto:aurelien.latouche@cnam.fr)>

**License** GPL-2

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2022-06-10 21:20:20 UTC

## R topics documented:

bce . . . . .	2
cdata . . . . .	2
center . . . . .	3
crrc . . . . .	3
crrs . . . . .	5
crrvvs . . . . .	8
print.crrs . . . . .	8
psh.test . . . . .	9
<b>Index</b>	<b>11</b>

---

bce

*Breast Cancer Data*

---

**Description**

Data Randomly Generated According To E1178 clinical trial

**Usage**

data(bce)

**Format**

A data frame with 200 observations and the following 6 variables.

trt Treatment: 0=Placebo, 1=Tamoxifen

time Event time

type Event type. 0=censored, 1=Breast Cancer recurrence , 2=Death without recurrence

nnodes Number of positive nodes

tsize Tumor size

age Age

**Examples**

data(bce)

---

cdata

*Clustered competing risks simulated data*

---

**Description**

sample of 200 observations

**Usage**

data(cdata)

**Format**

A data frame with 200 observations and the following 4 variables. Simulation is detailed on the paper Competing Risk Regression for clustered data. Zhou, Fine, Latouche, Labopin. 2011. In Press. Biostatistics.

ID Id of cluster, each cluster is of size 2

ftime Event time

fstatus Event type. 0=censored, 1 , 2

z a binary covariate with  $P(z=1)=0.5$

**Examples**

```
data(cdata)
```

---

```
center
```

---

```
Multicenter Bone Marrow transplantation data
```

---

**Description**

Random sub sample of 400 patients

**Usage**

```
data(center)
```

**Format**

A data frame with 400 observations and the following 5 variables.

id Id of transplantation center

ftime Event time

fstatus Event type. 0=censored, 1=Acute or Chronic GvHD , 2=Death free of GvHD

cells source of stem cells: peripheral blood vs bone marrow

fm female donor to male recipient match

**Examples**

```
data(center)
```

---

```
crrc
```

---

```
Competing Risks Regression for Clustered Data
```

---

**Description**

Regression modeling of subdistribution hazards for clustered right censored data. Failure times within the same cluster are dependent.

**Usage**

```
crrc(ftime, fstatus, cov1, cov2, tf, cluster,
      cengroup, failcode=1,
      cencode=0, subset,
      na.action=na.omit,
      gtol=1e-6, maxiter=10, init)
```

**Arguments**

<code>cluster</code>	Clustering covariate
<code>ftime</code>	vector of failure/censoring times
<code>fstatus</code>	vector with a unique code for each failure type and a separate code for censored observations
<code>cov1</code>	matrix (nobs x ncovs) of fixed covariates (either <code>cov1</code> , <code>cov2</code> , or both are required)
<code>cov2</code>	matrix of covariates that will be multiplied by functions of time; if used, often these covariates would also appear in <code>cov1</code> to give a prop hazards effect plus a time interaction
<code>tf</code>	functions of time. A function that takes a vector of times as an argument and returns a matrix whose <i>j</i> th column is the value of the time function corresponding to the <i>j</i> th column of <code>cov2</code> evaluated at the input time vector. At time <i>tk</i> , the model includes the term <code>cov2[, j]*tf(tk)[, j]</code> as a covariate.
<code>cengroup</code>	vector with different values for each group with a distinct censoring distribution (the censoring distribution is estimated separately within these groups). All data in one group, if missing.
<code>failcode</code>	code of <code>fstatus</code> that denotes the failure type of interest
<code>cencode</code>	code of <code>fstatus</code> that denotes censored observations
<code>subset</code>	a logical vector specifying a subset of cases to include in the analysis
<code>na.action</code>	a function specifying the action to take for any cases missing any of <code>ftime</code> , <code>fstatus</code> , <code>cov1</code> , <code>cov2</code> , <code>cengroup</code> , or <code>subset</code> .
<code>gtol</code>	iteration stops when a function of the gradient is < <code>gtol</code>
<code>maxiter</code>	maximum number of iterations in Newton algorithm (0 computes scores and var at <code>init</code> , but performs no iterations)
<code>init</code>	initial values of regression parameters (default=all 0)

**Details**

This method extends Fine-Gray proportional hazards model for subdistribution (1999) to accommodate situations where the failure times within a cluster might be correlated since the study subjects from the same cluster share common factors This model directly assesses the effect of covariates on the subdistribution of a particular type of failure in a competing risks setting.

**Value**

Returns a list of class `crr`, with components

<code>\$coef</code>	the estimated regression coefficients
<code>\$loglik</code>	log pseudo-likelihood evaluated at <code>coef</code>
<code>\$score</code>	derivatives of the log pseudo-likelihood evaluated at <code>coef</code>
<code>\$inf</code>	-second derivatives of the log pseudo-likelihood
<code>\$var</code>	estimated variance covariance matrix of <code>coef</code>
<code>\$res</code>	matrix of residuals

\$uftime	vector of unique failure times
\$bfitj	jumps in the Breslow-type estimate of the underlying sub-distribution cumulative hazard (used by predict.crr())
\$tfs	the tfs matrix (output of tf(), if used)
\$converged	TRUE if the iterative algorithm converged
\$call	The call to crr
\$n	The number of observations used in fitting the model
\$n.missing	The number of observations removed from the input data due to missing values
\$loglik.null	The value of the log pseudo-likelihood when all the coefficients are 0

### Author(s)

Bingqing Zhou, <bingqing.zhou@yale.edu>

### References

Zhou B, Fine J, Latouche A, Labopin M.(2012). Competing Risks Regression for Clustered data. *Biostatistics*. 13 (3): 371-383.

### See Also

cmprsk

### Examples

```
#library(cmprsk)
#crr(ftime=cdata$ftime, fstatus=cdata$fstatus, cov1=cdata$z)
# Simulated clustered data set
data(cdata)
crrc(ftime=cdata[,1], fstatus=cdata[,2],
cov1=cdata[,3],
cluster=cdata[,4])
```

---

crrs

*Competing Risks Regression for Stratified Data*

---

### Description

Regression modeling of subdistribution hazards for stratified right censored data

Two types of stratification are addressed : Regularly stratified: small number of large groups (strata) of subjects Highly stratified: large number of small groups (strata) of subjects

**Usage**

```
crrs(ftime, fstatus, cov1, cov2, strata,
     tf, failcode=1, cencode=0,
     ctype=1,
     subsets, na.action=na.omit,
     gtol=1e-6, maxiter=10, init)
```

**Arguments**

<code>strata</code>	stratification covariate
<code>ctype</code>	1 if estimating censoring dist within strata (regular stratification), 2 if estimating censoring dist across strata (highly stratification)
<code>ftime</code>	vector of failure/censoring times
<code>fstatus</code>	vector with a unique code for each failure type and a separate code for censored observations
<code>cov1</code>	matrix (nobs x ncovs) of fixed covariates (either <code>cov1</code> , <code>cov2</code> , or both are required)
<code>cov2</code>	matrix of covariates that will be multiplied by functions of time; if used, often these covariates would also appear in <code>cov1</code> to give a prop hazards effect plus a time interaction
<code>tf</code>	functions of time. A function that takes a vector of times as an argument and returns a matrix whose <i>j</i> th column is the value of the time function corresponding to the <i>j</i> th column of <code>cov2</code> evaluated at the input time vector. At time <code>tk</code> , the model includes the term <code>cov2[, j]*tf(tk)[, j]</code> as a covariate.
<code>failcode</code>	code of <code>fstatus</code> that denotes the failure type of interest
<code>cencode</code>	code of <code>fstatus</code> that denotes censored observations
<code>subsets</code>	a logical vector specifying a subset of cases to include in the analysis
<code>na.action</code>	a function specifying the action to take for any cases missing any of <code>ftime</code> , <code>fstatus</code> , <code>cov1</code> , <code>cov2</code> , <code>cengroup</code> , or <code>subset</code> .
<code>gtol</code>	iteration stops when a function of the gradient is < <code>gtol</code>
<code>maxiter</code>	maximum number of iterations in Newton algorithm (0 computes scores and var at <code>init</code> , but performs no iterations)
<code>init</code>	initial values of regression parameters (default=all 0)

**Details**

Fits the stratified extension of the Fine and Gray model (2011). This model directly assesses the effect of covariates on the subdistribution of a particular type of failure in a competing risks setting.

**Value**

Returns a list of class `crr`, with components (see `crr` for details)

<code>\$coef</code>	the estimated regression coefficients
<code>\$loglik</code>	log pseudo-likelihood evaluated at <code>coef</code>

<code>\$score</code>	derivitives of the log pseudo-likelihood evaluated at coef
<code>\$inf</code>	-second derivatives of the log pseudo-likelihood
<code>\$var</code>	estimated variance covariance matrix of coef
<code>\$res</code>	matrix of residuals
<code>\$uftime</code>	vector of unique failure times
<code>\$bfitj</code>	jumps in the Breslow-type estimate of the underlying sub-distribution cumulative hazard (used by <code>predict.crr()</code> )
<code>\$tfs</code>	the tfs matrix (output of <code>tf()</code> , if used)
<code>\$converged</code>	TRUE if the iterative algorithm converged
<code>\$call</code>	The call to <code>crr</code>
<code>\$n</code>	The number of observations used in fitting the model
<code>\$n.missing</code>	The number of observations removed from the input data due to missing values
<code>\$loglik.null</code>	The value of the log pseudo-likelihood when all the coefficients are 0

**Author(s)**

Bingqing Zhou, <bingqing.zhou@yale.edu>

**References**

Zhou B, Latouche A, Rocha V, Fine J. (2011). Competing Risks Regression for Stratified Data. *Biometrics*. 67(2):661-70.

**See Also**

`cmprsk`

**Examples**

```
##
#using fine and gray model
#crr(ftime=center$ftime, fstatus=center$fstatus,
#cov1=cbind(center$fm,center$cells))
#
# High Stratification: ctype=2
# Random sub-sample
data(center)
cov.test<-cbind(center$fm,center$cells)
crrs(ftime=center[,1],fstatus=center[,2],
cov1=cov.test,
strata=center$id,ctype=2)
```

---

crrvvs	<i>For internal use</i>
--------	-------------------------

---

**Description**

for internal use

**Author(s)**

Bingqing Zhou

---

print.crrs	<i>Print method for crrs output</i>
------------	-------------------------------------

---

**Description**

Prints call for crrs object

**Usage**

```
## S3 method for class 'crrs'  
print(x, ...)
```

**Arguments**

x	crr object (output from crrs())
...	additional arguments to print()

**Author(s)**

B. Zhou



---

psh.test	<i>Goodness-of-fit test for proportional subdistribution hazards model</i>
----------	--

---

### Description

This Goodness-of-fit test proposed a modified weighted Schoenfeld residuals to test the proportionality of subdistribution hazards for the Fine and Gray model

### Usage

```
psh.test(time, fstatus, z, D=c(1,1), tf=function(x) cbind(x,x^2), init)
```

### Arguments

time	vector of failure times
fstatus	failure status =0 if censored
z	covariates
D	components of z that are tested for time-varying effect
tf	functions of t for z being tested on the same location
init	initial values of regression parameters (default=all 0)

### Details

The proposed score test employs Schoenfeld residuals adapted to competing risks data. The form of the test is established assuming that the non-proportionality arises via time-dependent coefficients in the Fine-Gray model, similar to the test of Grambsch and Therneau.

### Value

Returns a data.frame with percentage of censored, cause 1, Test Statistic, d.f., p-value

### Author(s)

Bingqing Zhou, <bingqing.zhou@yale.edu>

### References

Zhou B, Fine JP, Laird, G. (2013). Goodness-of-fit test for proportional subdistribution hazards mode. *Statistics in Medicine*. In Press.

**Examples**

```
data(bce)
attach(bce)
lognodes <- log(nnodes)
Z1 <- cbind(lognodes, tsize/10, age, trt)
# trt = 0 if placebo, = 1 treatment
# testing for linear time varying effect of trt
psh.test(time=time, fstatus=type, z=Z1, D=c(0,0,0,1), tf=function(x) x)
# testing for quadratic time varying effect of trt
psh.test(time=time, fstatus=type, z=Z1, D=c(0,0,0,1), tf=function(x) x^2)
# testing for log time varying effect of trt
psh.test(time=time, fstatus=type, z=Z1, D=c(0,0,0,1),
tf=function(x) log(x))
# testing for both linear and quadratic time varying effect of trt
psh.test(time=time, fstatus=type, z=Z1,
D=matrix(c(0,0,0,1,0,0,0,1), 4,2), tf=function(x) cbind(x,x^2))
```

# Index

- \* **crrSC**
  - bce, 2
  - cdata, 2
  - center, 3
- \* **datasets**
  - bce, 2
  - cdata, 2
  - center, 3
- \* **survival**
  - crrc, 3
  - crrs, 5

bce, 2

cdata, 2

center, 3

crrc, 3

crrfs (crrvvs), 8

crrfsvs (crrvvs), 8

crrs, 5

crrvvs, 8

print.crrs, 8

psh.test, 9