Package 'SCCS'

October 16, 2026

Type Package

Title The Self-Controlled Case Series Method

Version 1.8

Date 2026-10-15

Description Various self-controlled case series models used to investigate associations between time-varying exposures such as vaccines or other drugs or non drug exposures and an adverse event can be fitted. Detailed information on the self-controlled case series method and its extensions with more examples can be found in Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018, ISBN: 978-1-4987-8159-6. Self-controlled Case Series studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press) and <https://sccs-studies.info>.

License GPL (>= 2)

Depends R (>= 2.14.0)

Imports survival, splines, corpcor, fda, R.methodsS3, gnm

LazyData yes

NeedsCompilation no

Author Yonas Ghebremichael Weldeselassie [aut, cre],
Heather Whitaker [aut],
Paddy Farrington [aut]

Maintainer Yonas Ghebremichael Weldeselassie <yonas.weldeselassie@open.ac.uk>

Repository CRAN

Date/Publication 2026-10-16 09:00:08 UTC

Contents

SCCS-package 2
addat 3
adidat 4
amdat 4

apdat . . . 5
autdat . . . 6
bpdat . . . 6
bupdat . . . 7
condat . . . 7
dtpdat . . . 8
eventdepenexp . . . 8
eventdepenobs . . . 10
febdat . . . 13
formatdata . . . 14
gbsdat . . . 16
gidat . . . 17
hibdat . . . 18
hipdat . . . 18
intdat . . . 19
integrateIspline . . . 19
itpdat . . . 20
lrtsccs . . . 21
midat . . . 22
nonparasccs . . . 22
nrtdat . . . 24
opvdat . . . 25
pmdat . . . 26
quantsccs . . . 26
rotdat . . . 27
rsvdat . . . 28
samplesize . . . 28
semisccs . . . 30
siddat . . . 32
simulatesccsdata . . . 33
smoothagesccs . . . 35
smoothexposccs . . . 37
standardsccs . . . 38

Index **44**

`SCCS-package` *The Self-Controlled Case Series (SCCS) Method*

Description

Fits the self-controlled case series model used to investigate the association between a time-varying exposures such as vaccines or other drugs and an adverse event. Some extensions of the SCCS method can be fitted with this package.

Note: Ages and times describing the start and end of observation, exposure times, and event times, should be expressed as integers. In most of the datasets, units of time are days; other choices may be appropriate according to context, but must be given as integers. Ages appearing as covariates need not be integers.

Details

Package:	SCCS
Type:	Package
Version:	1.7
Date:	2024-04-01
License:	"GPL-2"

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington

Maintainer: Yonas Ghebremichael-Weldeselassie <y.weldeselassie@warwick.ac.uk>

References

https://sccs-studies.info/index.html

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

`addat` *Data on NSAID and antidepressant exposure and first GI bleed*

Description

The data comprise ages in days at first gastro-intestinal (GI) bleed, treatments with non-steroidal anti-inflammatory drugs (NSAIDs), and treatments with antidepressants (SDs). There are 1000 simulated cases based on Tata et al (2005).

Usage

```
addat
```

Format

A data frame containing 3628 rows and 8 columns. The column names are 'case' (individual identifier), 'sta' (age on the first day of the observation period), 'end' (age on last day of the observation period), 'bleed' (age at first GI bleed), 'ns' (age at first day of NSAID treatment episode), 'endns' (age at last day of NSAID treatment episode), 'ad' (age at first day of AD treatment episode), 'endad' (age on last day of AD treatment episode). The data are in format 'stack' (see `dataformat` for details) with different exposure episodes on different rows.

References

Tata, L.J., Fortun, P. J., Hubbard, R. B., Smeeth, L., Hawkey, C. J., Smith, P., Whitaker, H. J., Farrington, C. P., Card, T. R. and West J. (2005). Does concurrent prescription of selective serotonin reuptake inhibitors and non-steroidal anti-inflammatory drugs substantially increase the risk of upper gastrointestinal bleeding? Alimentary Pharmacology and Therapeutics 22, 175–181.

`adidat` *Data on antidiabetics and fractures*

Description

The data comprise ages in days at first fracture and start of treatment with a thiazolidinedione antidiabetic. There are 2000 simulated cases based on Douglas et al (2009). Observation is curtailed at the end of treatment.

Usage

```
adidat
```

Format

A data frame containing 2000 rows and 6 columns. The column names are 'case' (individual identifier), 'frac' (age at fracture), 'sta' (age on the first day of the observation period), 'end' (age on last day of the observation period), 'adi' (age at first day of antidiabetic treatment), 'type' (fracture type: 1 for foot/ankle/wrist/hand, 2 for hip, 3 for spine).

References

Douglas, I. J., Evans, S. J., Pocock, S. and Smeeth L. (2009). The risk of fractures associated with thiazolidinediones: A self-controlled case-series study. PLoS Medicine 6 (9), e1000154.

`amdat` *Data on MMR and aseptic meningitis*

Description

The data comprise ages in days at measles, mumps and rubella (MMR) vaccination and hospital admission for aseptic meningitis. There are 10 admissions in 10 children.

Usage

```
amdat
```

Format

A data frame containing 10 rows and 5 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'am' (age at admission for aseptic meningitis, with a confirmed diagnosis of viral meningitis), and 'mmr' (age at mmr vaccination).

Source

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine 25, 1768–1797.

References

Miller, E., Goldacre, M., Pugh, S., Colville, A., Farrington, C.P., Flower, A., Nash, J., MacFarlane, L. and Tettmar, R. Risk of aseptic meningitis after measles, mumps and rubella vaccine in UK children (1993). The Lancet 341, 979–982.

---

apdat *Data on antipsychotics and stroke*

Description

The data comprise ages in days at first stroke and treatments with antipsychotics. There are 2000 simulated cases based on Douglas and Smeeth (2008), including 1500 without dementia and 500 with dementia.

Usage

```
apdat
```

Format

A data frame containing 11792 rows and 8 columns. The column names are 'case' (individual identifier), 'sta' (age on the first day of the observation period), 'end' (age on last day of the observation period), 'stro' (age at first stroke), 'ap' (age at first day of antipsychotic treatment episode), 'endap' (age at last day of antipsychotic treatment episode), 'cen' (1 if observation ended before the end of the study, 0 otherwise), 'dem' (0 if the patient does not have dementia, 1 if the patient has dementia). The data are in format 'stack' with different exposure episodes on different rows.

References

Douglas, I. J. and Smeeth L. (2008). Exposure to antipsychotics and risk of stroke: Self-controlled case series study. British Medical Journal 337, a1227.

autdat *Data on MMR vaccine and autism*

Description

The data comprise ages in days at measles mumps rubella (MMR) vaccination and autism diagnosis. There are 350 simulated cases based on Taylor et al (1999).

Usage

```
autdat
```

Format

A data frame containing 350 rows and 5 columns. The column names are 'case' (individual identifier), 'diag' (age at autism diagnosis), 'sta' (age on the first day of the observation period), 'end' (age on last day of the observation period), 'mmr' (age at MMR vaccination.

References

Taylor, B., Miller, E., Farrington, C. P., Petropoulos, M.C., Favot-Mayaud, I., Li, J., and Waight, P. A. (1999). Autism and measles, mumps and rubella vaccine: No epidemiological evidence for a causal association. Lancet 353, 2026-2029.

bpdat *Data on blood pressure and headaches*

Description

The data comprise systolic and diastolic blood pressures taken twice a day for seven days, and start of headache in the next period. There are 71 headaches in 64 cases.

Usage

```
bpdat
```

Format

A data frame containing 896 rows and 6 columns. The column names are 'case' (individual identifier), 'dow' (day of week, coded 1 for Monday,...,7 for Sunday), 'time' (time of reading, coded 1 for am, 2 for pm), 'sys' (systolic blood pressure), 'dia' (diastolic blood pressure), 'head' (1 if a headache started in the period up to the next blood pressure reading, or within 12 hours of the last reading).

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

bupdat *Data on bupropion and sudden death*

Description

The data comprise simulated data on bupropion and sudden death. There are 121 cases. The start of observation coincides with age at first bupropion prescription; the nominal end of observation is age on 11 November 2003 (day 1136). Ages are in days.

Usage

```
bupdat
```

Format

A data frame containing 121 rows and 4 columns. The column names are 'case' (individual identifier), 'date' (date of first bupropion prescription, day 0 = 1 October 2000), 'bup' (age at first bupropion prescription), 'death' (age at sudden death).

References

Hubbard R., Lewis S., West J., Smith C., Godfrey C., Smeeth L., Farrington P. and Britton J. (2005). Bupropion and the risk of sudden death: a self-controlled case series analysis using The Health Improvement Network. Thorax 60, 848-850.

condat *Data on DTP and convulsions*

Description

The data comprise ages in days at measles, mumps and rubella (MMR) vaccination, Haemophilus influenzae type b (Hib) booster or catch-up vaccination, and febrile convulsion. There are 2435 convulsions in 2201 children. The ages have been jittered.

Usage

```
condat
```

Format

A data frame containing 2435 rows and 11 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'conv' (age at convulsion), 'hib' (age at Hib booster/catch-up vaccination), 'mmr' (age at MMR vaccination), 'sex' (1 for males, 2 for females), 'gap' (days from convulsion to next convulsion within the same case, or to end of observation), 'cen' (0 if last admission for this case, 1 otherwise), 'rec' (within-case event number), 'ngrp' (1 if case has a unique event, 2 if case has 2+ events).

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

`dtpdat` *Data on DTP and convulsions*

Description

The data comprise ages in days at diphteria, tetanus and pertussis (DTP) vaccination and febrile convulsion. There are 1379 convulsions in 1214 children. The ages have been jittered.

Usage

```
dtpdat
```

Format

A data frame containing 1379 rows and 8 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'conv' (age at convulsion), 'dtp' (age at first dose of DTP), 'dtpd2' (age at second dose of DTP), 'dtpd3' (age at third dose of DTP), 'sex' (1 for males, 2 for females).

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

`eventdepenexp` *SCCS with event-dependent exposure*

Description

One of the assumptions of the self-controlled case series model is that occurence of an event does not affect subsequent exposure. This function fits the modified SCCS model when the assumption is not satisfied, see Farrington et al (2009). This modified method assumes that no exposure is possible following a unique event. It requires that exposure is of a fixed finite length and that the end of observation that would have applied in the absence of the event is known.

Usage

```
eventdepenexp(indiv, astart, aend, aevent, adrug, aedrug, expogrp=0,
              sameexpopar=T,agegrp=NULL, dataformat="stack", verbose=F,
              tolerance=1e-8,itermax=100, data)
```

Arguments

indiv	a vector of individual identifiers of cases
astart	a vector of ages at which the observation periods start
aend	a vector of ages at end of observation periods, that would have applied in the absence of the event
aevent	a vector of ages at event (one event per case)
adrug	a vector of ages at which exposure starts or a matrix if there are multiple episodes of the same exposure type (dataformat multi). Multiple exposures of the same exposure type can be recorded as multiple rows (dataformat stack). In this method only one exposure type can be included unlike the standardsccs where multiple exposure types can be analayzed in the same model
aedrug	a vector of ages at which exposure-related risk ends or a matrix if there are multiple episodes of the same exposure type. The dimension of aedrug should be equal to the dimension of adrug, that is aedrug should be given for each column in adrug
expogrp	a vector of days to the start of exposure-related risk, counted from adrug. E.g if the risk period is [adrug+c,aedrug], use expogrp = list(c) or expogrp = c. The DEFAULT is a expogrp= 0 where the exposure-related risk period is [adrug, aedrug].
sameexpopar	a logical value. If TRUE (the default) no dose effect is assumed: the same exposure parameters are used for multiple doses/episodes of the same exposure type presented in dataformat 'multi'. If FALSE different relative incidences are estimated for different doses/episodes of the same exposure.
agegrp	a vector of cut points for the age groups where each value represents the start of an age catagory. The first element in the vector is the start of the second age group. The first age group starts at the minimum of astart, the start of observation period. The default is NULL (i.e no age effects included).
dataformat	the way the input data are assembled. It accepts "multi" or "stack" (the default), where "multi" refers to a data assembled with one row representing one event and "stack" refers to a data frame where repeated exposures of the same type are stacked in one column. In the "multi" dataformat different episodes of the same exposure type are recorded as separate columns in the dataframe.
verbose	a logical value indicating whether information about the iterations should be printed. Default is FALSE
tolerance	the convergence tolerance when estimating the parameters. Defaults to 1e-8.
itermax	maximum number of iterations. 100 is the default.
data	a data frame containing the input data. The data should be in 'stack' or 'multi' (see dataformat).

Details

This model fits a SCCS model with event-dependent exposures.

Value

Relative incidence estimates along with their 95% confidence intervals.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington, C. P., Whitaker H.J., and Hocine M.N. (2009). Case series analysis for censored, perturbed or curtailed post-event exposures. Biotatistics, 10(1), 3-16.

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

semisccs, standardsccs

Examples

```
library(SCCS)

# Analysis of rotavirus vaccination and intussusception data
# Model 1: Three doses of the same vaccine exposure OPV (OPV, OPV2 and opv3),
# only one risk period [adrug, aedrug]


   rot.mod1 <- eventdepenexp(indiv=case, astart=sta, aend=end,
                         aevent=intus, adrug=cbind(rv,rvd2),
                         expogrp=1,aedrug=cbind(rv+21,rvd2+21),
                         agegrp=seq(56,168,14), dataformat="multi", data=rotdat)

   rot.mod1

# Model 2: Two doses with two riks periods, 1-7 and 8-21


   rot.mod2 <- eventdepenexp(indiv=case, astart=sta, aend=end,
                            aevent=intus, adrug=cbind(rv,rvd2),
                            aedrug=cbind(rv+21,rvd2+21), expogrp=c(1,8),
                            agegrp=seq(56,168,14), dataformat="multi",
                            data=rotdat)

   rot.mod2
```

---

eventdepenobs *SCCS with event-dependent observation periods*

Description

One of the assumptions of the self-controlled case series models is that the observation period for each individual is independent of event times. If an event increases the risk of death, such as myocardial infraction or stroke, this assumption is violated. This function fits the modified SCCS model when the assumption is not satisfied i.e ages at end of observation periods might depend on age at event as outlined in Farrington et al (2011).

Usage

```
eventdepenobs(formula, indiv, astart, aend, aevent, adrug, aedrug, censor,
              expogrp = list(), washout = list(), sameexpopar = list(),
              agegrp = NULL, dataformat="stack", covariates=NULL,
              regress=F, initval=rep(0.1, 7), data)
```

Arguments

formula	a model formula. The dependent variable should always be "event" e.g. event ~ itp. If age effects are included in the model, the word 'age' must be used in the formula, e.g event ~ itp + age.
indiv	a vector of individual identifiers of cases
astart	a vector of ages at which the observation periods start
aend	a vector of ages at end of observation periods
aevent	a vector of ages at event (one event per case)
adrug	a list of vectors of ages at start of exposures or a list of matrices if the exposures have multiple episodes (dataformat multi). Multiple exposures of the same type can be recorded as multiple rows (dataformat stack). One list item per exposure type.
aedrug	a list of vectors of ages at which exposure-related risk ends or a list of matrices if there are multiple episodes (repeat exposures in different columns) of the same exposure type. The dimension of each item of aedrug has to be equal to that of adrug, that is aedrug should be given for each exposure in adrug.
censor	a vector of indicators for whether an observation periods were censored (1 = observation period ended early, 0 = fully observed).
expogrp	list of vectors of days to the start of exposure-related risk, counted from adrug. E.g if the risk period is [adrug+c,aedrug], use expogrp = list(c) or expogrp = c. For multiple exposure types expogrp is a list of vectors having the same length as list adrug. The DEFAULT is a list of zeros where the exposure-related risk periods are [adrug, aedrug].
washout	list of vectors with days on start of washout periods counted from aedrug, the number of vectors in the list is equal to the number of exposures or the length of list of adrug. The default is NULL, no washout periods. The order of the list corresponds to the order of exposures in adrug.
sameexpopar	a vector of logical values. If TRUE (the default) no dose effect is assumed, the same exposure parameters are used for multiple doses/episodes of the same exposure type presented in dataformat 'multi'. If FALSE different relative

incidences are estimated for different doses/episodes of the same exposure type. The length of the vector is equal to the length of the list adrug.

`agegrp` a vector of cut points for the age groups where each value represents the start of an age catagory. The first element in the vector is the start of the second age group. The first age group starts at the minimum of `astart`, the start of observation period. The defaults is NULL (i.e no age effects included).

`dataformat` the way the input data are assembled. It accepts "multi" or "stack" (the default), where "multi" refers to a data assembled with one row representing one event and "stack" refers to a data frame where repeated exposures of the same exposure type are stacked in one column. In the "multi" dataformat different episodes of the same exposure type are recorded as separate columns in the dataframe.

`covariates` list of covariates believed to affect the age at censoring (age at end of observation period) (e.g. covariates = gender).

`regress` logical, `regress=T` indicates that the parameters of the weight functions are regressed againt age at event or age at start of observation. The default is `regress=F`

`initval` a vector of intial values used in fitting the weight functions. These are given in the order of: 1. Log mean of the exponential component 2. Intercept of the EG/EW log mean function 3. Intercept of the EG/EW log shape function 4. Intercept of the logit mixing probability function 5. Regression parameter of the G/W log mean functions, if regress = T 6. Regression parameter of the G/W log shape function, if regress=T 7. Regression parameter of the G/W logit mixing probability function, if regress=T. When `regress=F` only the first 4 are used. The default `initval` values are 0.1.

`data` a data frame containing the input data. The data should be in 'stack' or 'multi' (see `dataformat`).

Details

This model is suitable when the event increases the risk of death, such as myocardial infraction (MI) or stroke. It is not suitable when the event itself is death. Four models are fitted to the interval between the age at end of observation and the event date, these are detailed in section 5.4 of Farrington et al (2011). The model with the lowest AIC is selected, and used to estimate weights that replace interval lengths in the model formula. This modification allows unbiased estimates of the exposure effect to be estimated, while age effects take on a different interpretation as they include the thinning effect of censoring.

Value

`summary` exposure related relative incidence estimates along with their 95% confidence intervals, age related relative incidence estimates and estimates of interactions with covariates if there are any.

`modelfit` model fit of the 4 different weight functions and their AIC values.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington, C. P., Anaya-Izquierdo, A., Whitaker, H. J., Hocine, M.N., Douglas, I., and Smeeth, L. (2011). Self-Controlled case series analysis With event-Dependent observation periods. Journal of the American Statistical Association 106 (494), 417–426.

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

Examples

```
library(SCCS)

# Nicotine replacement therapy and myocardial infarction (MI)
# With no age effect included


nrt.mod <- eventdepenobs(event~nrt, indiv=case, astart=nrt,
            aend=act, aevent=mi, adrug=nrt, aedrug=nrt+28,
            censor=cen, expogrp=c(0,8,15,22), agegrp=NULL,
            data=nrtdat)


# Respiratory tract infections and MI
# Age effect included
# intial values provided and there are two risk periods

uni <- (1-duplicated(midat$case))
ageq <- floor(quantile(midat$mi[uni==1], seq(0.1,0.9,0.1), names=FALSE))
                  # age groups

mi.mod <- eventdepenobs(event~rti+age, indiv=case, astart=sta,
                        aend=end, aevent=mi, adrug=rti, aedrug=rti+14,
                        expogrp=c(0,8), agegrp=ageq, censor=cen, data=midat,
                        initval=rep(1.1,4))

mi.mod
```

---

febdat *Data on multitype convulsions and MMR*

Description

The data comprise ages in days at MMR vaccine and convusion in children aged 366 to 730 days of age. The convulsions are two types: febrile or non-febrile. There are 988 events in 894 cases.

Usage

```
febdat
```

Format

A data frame containing 988 rows and 7 columns. The column names are 'case' (individual identifier), 'conv' (age at convulsion), 'sta' (age at start of observation period), 'end' (age at end of observation period), 'mmr' (age at MMR vaccine), 'sex' (coded 1 for males, 2 for females), 'type' (coded 1 for non-febrile convulsion, 2 for febrile convulsion).

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

`formatdata` *Formatting data*

Description

Reformats the data based on age and/or season and exposure groups prior to fitting SCCS model.

Usage

```
formatdata(indiv, astart, aend, aevent, adrug, aedrug, expogrp = list(),
           washout = list(), sameexpopar = list(), agegrp = NULL,
           seasongrp=NULL, dob=NULL, cov = cbind(), dataformat="stack", data)
```

Arguments

`indiv`	a vector of individual identifiers of cases
`astart`	a vector of ages at which the observation periods start
`aend`	a vector of ages at end of observation periods
`aevent`	a vector of ages at event, an individual can experience multiple events
`adrug`	a list of vectors of ages at start of exposures or a list of matrices if the exposures have multiple episodes (`dataformat` multi). Multiple exposures of the same type can be recorded as multiple rows (`dataformat` stack). One list item per exposure type.
`aedrug`	a list of vectors of ages at which exposure-related risk ends or a list of matrices if there are multiple episodes (repeat exposures in different columns) of the same exposure type. The dimension of each item of `aedrug` has to be equal to that of `adrug`, that is `aedrug` should be given for each exposure in `adrug`.
`expogrp`	list of vectors of days to the start of exposure-related risk, counted from `adrug`. E.g if the risk period is [`adrug`+c,`aedrug`], use expogrp = list(c) or expogrp = c. For multiple exposure types `expogrp` is a list of vectors equal to the length of the list of `adrug`. The DEFAULT is a list of zeros where the exposure-related risk periods are [`adrug`, `aedrug`].

washout	list of vectors with days on start of washout periods counted from aedrug, the number of vectors in the list is equal to the number of exposures or the length of list of adrug. The default is NULL, no washout periods. The order of the list corresponds to the order of exposures in adrug.
sameexpopar	a vector of logical values. If TRUE (the default) no dose effect is assumed, the same exposure parameters are used for multiple doses/episodes of the same exposure type presented in dataformat 'multi'. If FALSE different relative incidences are estimated for different doses/episodes of the same exposure. The length of the vector is equal to the length of adrug. The order in which the elements of the vector are put corrosponds to the order of exposures adrug.
agegrp	a vector of cut points of the age groups where each value represents the start of an age catagory. The first element in the vector is the start of the second age group. The first age group starts at astart, the start of observation period. The defaults is NULL (i.e no age effects included).
seasongrp	a vector of cut points for seasonal effects. The values should be given in ddmm format, representing the first days of each season group. The seasonal effect is a factor, the reference level being the time interval starting at the earliest date in seasongrp. The default is NULL where no seasonal effects are included in the model.
dob	a vector of birth dates of the cases, in ddmmyyyy format. They are used if seasonal effects are included in the model. The default dob is NULL but is required if seasongrp is not NULL.
cov	a vector (or a matrix if there are multiple) of fixed covariates. The default is NULL where no covariates are included.
dataformat	the way the input data are assembled. It accepts "multi" or "stack" (the default), where "multi" refers to a data assembled with one row representing one event and "stack" refers to a data frame where repeated exposures of the same type are stack in one column. In the "multi" dataformat different episodes of the same exposure type are recorded as separate columns in the dataframe.
data	a data frame containing the input data. The data should be in 'stack' or 'multi' (see dataformat).

Value

a data frame containing the following columns:

indivL	an identfier for each individual event.
event	indicator for presence of an event within an interval. "1" where an event occured, "0" otherwise.
age	factor for age groups.
Season	a factor for season if seasongrp is specified.
exposures	factors for exposure status of each exposure type. "0" for baseline/control periods, "1" for the first risk period. "1" for subsequent exposure risk periods if sameexpopar=TRUE, or increasing factor levels for each subsequent exposure if sameexpopar = FALSE. Indicators for washout periods (if there are any) are also included here. The column names of these factors are the same as the column names of the exposures in adrug.

`interval` length of interval. Needed for offsets within the model.

There are also columns for eventday (day of adverse event), lower (day a period starts), upper (day a period ends), indiv (original individual indentifier), aevent, astart, aend and any covariates included in `cov`.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine 25, 1768–1797.

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

Examples

```
# MMR vaccine and ITP data

# A single exposure with three risk periods and no age groups included

 itp.dat1 <- formatdata(indiv=case, astart=sta, aend=end,
                      aevent=itp, adrug=mmr, aedrug=mmr+42,
                      expogrp=c(0,15,29),
                      data=itpdat)

 itp.dat1

# A single exposure with three risk periods and six age groups

 itp.dat2 <- formatdata(indiv=case, astart=sta, aend=end,
                      aevent=itp, adrug=mmr, aedrug=mmr+42,
                      expogrp=c(0,15,29), agegrp=c(427,488,549,610,671),
                      data=itpdat)
 itp.dat2
```

---

gbsdat *Data on influenza vaccine and GBS*

Description

The data comprise days (day 1 is 1st October 2010) at seasonal influenza vaccination and at onset of Guillain-Barre Syndrome (GBS) in Italy, gathered in the 2010-2011 influenza season. There are 174 cases from Galeotti (2013). Times have been jittered.

Usage

```
gbsdat
```

Format

A data frame containing 174 rows and 6 columns. The column names are 'case' (individual identifier), 'sta' (the first day of the observation period), 'end' (the last day of the observation period), 'gbs' (day of GBS onset), 'flu' (day of influenza vaccination), 'sage' (age in years on day 1, 1st October 2010).

References

Galeotti, F., M. Massari, R. D'Alessandro, E. Beghi, A. Chio, G. Logroscino, G. Filippini, M. D. Benedetti, M. Pugliatti, C. Santiccio, and R. Raschetti (2013). Risk of Guillain-Barre syndrome after 2010-2011 influenza vaccination. European Journal of Epidemiology 28 (5), 433-444.

---

gidat *Data on NSAID and first GI bleed*

Description

The data comprise ages in days at first gastro-intestinal (GI) bleed and treatments with non-steroidal anti-inflammatory drugs (NSAIDs). There are 838 simulated cases based on Tata et al (2005).

Usage

```
gidat
```

Format

A data frame containing 2920 rows and 6 columns. The column names are 'case' (individual identifier), 'sta' (age on the first day of the observation period), 'end' (age on last day of the observation period), 'bleed' (age at first GI bleed), 'ns' (age at first day of NSAID treatment episode), 'endns' (age at last day of NSAID treatment episode). The data are in format 'stack' with different exposure episodes on different rows.

References

Tata, L. J., Fortun, P. J., Hubbard, R. B., Smeeth, L., Hawkey, C. J., Smith, P., Whitaker, H. J., Farrington, C. P., Card, T. R., and West J. (2005). Does concurrent prescription of selective serotonin reuptake inhibitors and non-steroidal anti-inammatory drugs substantially increase the risk of upper gastrointestinal bleeding? Alimentary Pharmacology and Therapeutics 22, 175–181.

hibdat *Data on DTP, Hib and convulsions*

Description

The data comprise ages at diphteria, tetanus and pertussis (DTP) vaccination, Haemophilus influenzae type b (Hib) vaccination, and febrile convulsion. There are 1378 convulsions in 1213 children. The ages have been jittered.

Usage

```
hibdat
```

Format

A data frame containing 1378 rows and 11 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'conv' (age at convulsion), 'dtp' (age at first dose of DTP), 'dtpd2' (age at second dose of DTP), 'dtpd3' (age at third dose of DTP), 'hib' (age at first dose of Hib), 'hibd2' (age at second dose of Hib), 'hibd3' (age at third dose of hib), 'sex' (1 for males, 2 for females).

hipdat *Data on antidepressants and hip fracture*

Description

The data comprise ages in days at first treatment with antidepressant and first hip fracture. There are 1000 simulated cases based on Hubbard et al (2003).

Usage

```
hipdat
```

Format

A data frame containing 1000 rows and 6 columns. The column names are 'case' (individual identifier), 'frac' (age at first hip fracture), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'ad' (age at start of first treatment with antidepressant), 'endad' (age at end of first antidepressant treatment).

References

Hubbard, R., Farrington, C.p., Smith, C., Smeeth, L., and Tattersfield A. (2003). Exposure to tricyclic and selective serotonin reuptake inhibitor antidepressants and the risk of hip fracture. American Journal of Epidemiology 158 (1), 77–84.

intdat *Data on intussusception and OPV*

Description

The data comprise ages in days at intussusception and oral polio vaccine (OPV) from Cuba. There are 273 intussusceptions in 273 children. The ages have been jittered.

Usage

```
intdat
```

Format

A data frame containing 273 rows and 7 columns. The column names are 'case' (individual identifier), 'intus' (age at intussusception), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'opv' (age at first dose of OPV), 'opvd2' (age at second dose of OPV), 'dob' (date of birth, in ddmmyyyy format).

References

Sardinas, M.A.G., Cardenas, A.z., Marie, G.c., Santiago, M.A., Sanchez, M.V., and Farrington C. P. (2001). Lack of association between intussusception and oral polio vaccine in Cuban children. European Journal of Epidemiology 17 (8), 783-787.

integrateIspline *Integral of I-splines*

Description

Evaluates design matrix for integrals of I-splines and integrals of the integrals. The function evaluates first, second and third integrals of I-splines.

Usage

```
integrateIspline(x, knots1, m, int)
```

Arguments

x	a numeric vector of values at which to evaluate the integrals of I-spline functions
knots1	a numeric vector of interior knot positions with non-decreasing values
m	a positive integer giving the order of the spline function.
int	a positive integer (1, 2, or 3) for first, second or third integral of an I-spline function.

Value

A matrix with length of (x) rows and length of (knots1) - m columns.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Ghebremichael-Weldeselassie Y. (2014). Smooth risk functions for self-cotrolled case series models. PhD thesis, The Open University.

Ghebremichael-Weldeselassie, Y., Whitaker, H. J., Farrington, C. P. (2015). Spline-based self controlled case series method. Statistics in Medicine 33:639-649.

`itpdat` *Data on MMR and ITP*

Description

The data comprise ages in days at measles, mumps and rubella (MMR) vaccination and hospital admission for idiopathic thrombocytopaenic purpura (ITP). There are 44 admissions in 35 children.

Usage

```
itpdat
```

Format

A data frame containing 44 rows and 9 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'itp' (age at admission for ITP), 'mmr' (age at mmr vaccination), 'sex' (1 for males, 2 for females), 'gap' (days from ITP admission to next ITP admission within the same case, or to end of observation), 'cen' (0 if last admission for this case, 1 otherwise), 'rec' (within-case event number).

Source

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine 25, 1768–1797.

References

Miller, E., Waight, P., Farrington, P., Andrews, N., Stowe, J., and Taylor B. (2001). Idiopathic thrombocytopenic purpura and MMR vaccine. Archives of Disease in Childhood 84, 227–229.

`lrtsccs` *Likelihood ratio test for SCCS models*

Description

The function performs the likelihood ratio test for SCCS models that are nested (up to combining of multinomial categories).

Usage

```
lrtsccs(model1, model2)
```

Arguments

`model1`	an object fitted by the SCCS method e.g `standardsccs` to be compared with `model2`.
`model2`	an object fitted by the SCCS method e.g `standardsccs` to be compared with `model1`.

Value

likelihood ratio test statistic, degrees of freedom and p-value.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

Examples

```
itp.mod1 <- standardsccs(event~mmr+age, indiv=case, astart=sta,aend=end,
                         aevent=itp, adrug=mmr, aedrug=mmr+42, expogrp=c(0,15,29),
                         agegrp=c(427,488,549,610,671), data=itpdat)


itp.mod2 <- standardsccs(event~age, indiv=case, astart=sta,
                        aend=end, aevent=itp, adrug=mmr, aedrug=mmr+42,
                        expogrp=c(0,15,29), agegrp=c(427,488,549,610,671),
                        data=itpdat)

itp.mod3 <- standardsccs(event~mmr + age, indiv=case, astart=sta,
                         aend=end, aevent=itp, adrug=mmr, aedrug=mmr+42,
                         agegrp=c(427,488,549,610,671), data=itpdat)
```

```
# Compare itp.mod1 a model with both age and exposure (mmr) and itpmod2 a model
# with only age effect

   lrtsccs(itp.mod1,itp.mod2)

# Compare itp.mod1 a model with both age and 3 exposure categories and itpmod3
# a model with age and only one exposure category

   lrtsccs(itp.mod3,itp.mod1)  # order of the objects doesn't matter
```

`midat` *Data on respiratory tract infections and myocardial infarction*

Description

These simulated data comprise ages in days at respiratory tract infection and first myocardial infarction in 1000 cases.

Usage

```
midat
```

Format

A data frame containing 2187 rows and 6 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'mi' (age at first myocardial infarction), 'rti' (age at respiratory tract infection), 'cen' (the censoring indicator, equal to 1 if the observation period is censored, 0 if not). The data are in format "stack".

References

Smeeth L., Thomas, S. L., Hall, A. J., Hubbard, R., Farrington, P., Vallance, P. (2004). Risk of myocardial infarction and stroke after acute infection or vaccination. New England Journal of Medicine 351, 2611-2618.

`nonparasccs` *Spline-based non parametric SCCS method*

Description

Fits a spline-based non parametric SCCS model where both the exposure related relative incidence and age related relative incidence functions are represented by spline functions; that is, linear combinations of M-splines.

Usage

```
nonparasccs(indiv, astart, aend, aevent, adrug, aedrug, kn1=12, kn2=12,
            sp1=NULL, sp2=NULL, data)
```

Arguments

indiv	a vector of individual identifiers of cases.
astart	a vector of ages at start of observation periods.
aend	a vector of ages at end of observation periods.
aevent	a vector of ages at event, an individual can experience multiple events.
adrug	a vector of ages at which exposure related risk period starts.
aedrug	a vector of ages at which exposure-related risk ends.
kn1	an integer >= 5 representing the number of interior knots used to define the M-spline basis functions which are related to the age specific relative incidence function, usually between 8 and 12 knots is sufficient. It defaults to 12 knots.
kn2	a an integer >= 5 representing the number of interior knots used to define the M-spline basis functions which are related to the exposure specific relative incidence function, usually between 8 and 12 knots is sufficient. The default value is 12.
sp1	smoothing parameter value for age related relative incidence function. It defaults to "NULL" where the smoothing parameter is obtained automatically using an approximate cross-validation method. The value of "sp1" must be a number greater or equal to 0.
sp2	smoothing parameter value for exposure related relative incidence function. It defaults to "NULL" where the smoothing paramter is obtained automatically using an approximate cross-validation method. The value of "sp1" must be a number greater or equal to 0.
data	A data frame containing the input data.

Details

The smoothing parameters for the age and exposure related relative incidence functions are chosen using a cross-validation method. To visualize the exposure-related relative incidence function, use the plot function.

Value

Relative incidence estimates along with their 95% confidence intervals.

estimates	exposure related relative incidence estimates at each point of time since start of exposure until the maximum difference between the start and end of exposure.
timesinceexposure	time units since the start of exposure.
lci	lower confidence limits of the exposure related relative incidence estimates.
uci	upper confidence limits of the exposure related relative incidence estimates.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Ghebremichael-Weldeselassie, Y., Whitaker, H. J., Farrington, C. P. (2016). Flexible modelling of vaccine effects in self-controlled case series models. Biometrical Journal, 58(3):607-622.

Ghebremichael-Weldeselassie, Y., Whitaker, H. J., Farrington, C. P. (2017). Spline-based self controlled case series method. Statistics in Medicine 33:639-649.

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

smoothagesccs, smoothexposccs

Examples

```
# ITP and MMR data

  itp.mod <- nonparasccs(indiv=case, astart=sta, aend=end,
                         aevent=itp, adrug=mmr, aedrug=mmr+42, sp1=28000, sp2=1200,
                         data=itpdat)

  itp.mod

# Plot the exposure and age related relative incidence functions

  plot(itp.mod)
```

---

nrtdat *Data on NRT and MI*

Description

The data comprise ages in days at first treatment with nicotine replacement therapy (NRT) and first subsequent myocardial infarction (MI). There are 141 simulated cases based on Tata et al (2005).

Usage

```
nrtdat
```

Format

A data frame containing 141 rows and 7 columns. The column names are 'case' (individual identifier), 'nrt' (age at initiation of NRT treatment), 'mi' (age at first subsequent MI), 'end' (age at end of nominal observation period, nrt + 365), 'act' (age at earliest of end and actual end of observation period), 'cage' (centred age, in years, at NRT), 'cen' (1 if act = end, 0 if act < end).

References

Hubbard, R., Lewis, S., Smith, C., Godfrey, C., Smeeth, L., Farrington, P., and Britton J. (2005). Use of nicotine replacement therapy and the risk of acute myocardial infarction, stroke and death. Tobacco Control 14, 416-421.

---

opvdat *Data on OPV and intussusception*

Description

The data comprise ages in days at oral polio vaccine (OPV) and first hospital admission for intussusception in children between the ages of 27 and 365 days. There are 207 cases.

Usage

```
opvdat
```

Format

A data frame containing 207 rows and 8 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'intus' (age at admission for intussusception), 'opv' (age at first dose of OPV), 'opvd2' (age at second dose of OPV), 'opvd3' (age at third dose of OPV), 'sex' (1 for males, 2 for females).

Source

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine 25, 1768–1797.

References

Andrews N., Miller, E., Waight, P., Farrington, P., Crowcroft, N., Stowe, J., and Taylor B. (2002). Does oral polio vaccine cause intussusception in infants? Evidence from a sequence of three self-controlled case series studies in the United Kingdom. European Journal of Epidemiology 17, 701–706.

pmdat *Data on asthma admissions and air pollution*

Description

The data are a daily time series of hospital admissions for asthma in Nottingham, with levels (in micrograms per cubic metre) of particulate matter less than 10 micrometres in diameter (PM10 levels).

Usage

```
pmdat
```

Format

A data frame containing 2922 rows and 3 columns. The column names are 'day' (day of observation, 1 to 2922), 'asma' (number of asthma admissions on that day), and 'pm10' (PM10 level on that day). The data are in time series format, for SCCS analysis using generalised linear models.

References

Farrington, C. P. and Whitaker H. J. (2006). Semiparametric analysis of case series data (with discussion). Journal of the Royal Statistical Society, Series C 55 (5), 553-594.

quantsccs *Quantitative exposures in self controlled case series method*

Description

This function fits the starndard SCCS model where the exposures are measured on a continuous scale.

Usage

```
quantsccs(formula, indiv, event, data)
```

Arguments

formula	model formula. The dependent variale should always be "event" e.g. event ~ expo, where expo is an exposure measured at each time unit.
indiv	a vector of individual identifiers of cases
event	number of events occuring at each time unit.
data	a data frame containing the input data. Data are assembled one line per time unit of observation.

Details

In this method exposures are measured at successive time points within the observation period for each case. And number of events experienced by each case at each time point are recorded.

Value

Relative incidence estimates along with their 95% confidence intervals.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

semisccs

Examples

```
# Headaches and blood pressure data. sys and dia (continuoous exposures) are systolic and
# diastolic blood pressure measurements respectively

 bp.mod <- quantsccs(event~sys+dia, indiv=case, event=head,
                     data=bpdat)

 bp.mod
```

`rotdat` *Data on Rotavirus vaccine and intussusception*

Description

The data comprise ages in days at rotavirus vaccine (RV) and first symptoms of intussusception in children between the ages of 42 and 183 days. There are 566 cases. Ages have been jittered.

Usage

```
rotdat
```

Format

A data frame containing 566 rows and 6 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the observation period), 'intus' (age at first symptom of intussusception), 'rv' (age at first dose of RV), 'rvd2' (age at second dose of RV).

References

Stowe J., Andrews, N., Ladhani, S., and Miller E. (2016). The risk of intussusception following monovalent rotavirus vaccination in England: A self-controlled case-series evaluation. Vaccine 34, 3684-3689.

`rsvdat` *Data on RSV and ambient temperature*

Description

The data are a weekly time series of counts of respiratory syncytial virus (RSV) isolates from England and Wales and average temperatures (in degrees Celsius) in Central England.

Usage

```
rsvdat
```

Format

A data frame containing 364 rows and 5 columns. The column names are 'year' (year of report, 1996 to 2003), 'week' (week of the year, numbered 1 to 52), 'win' (4-week window, numbered 1 to 91), 'temp' (average daily temperature for the week prior to the current week), and 'rsv' (count of RSV isolates for the current week). The data are in time series format, for SCCS analysis using generalised linear models.

References

Whitaker, H. J., Hocine, N. and Farrington C. P. (2007). On case-crossover methods for environmental time series data. Environmetrics 18 (2), 157-171.

`samplesize` *Sample size calculation in SCCS*

Description

The function calculates the sample size required for an SCCS analysis.

Usage

```
samplesize(eexpo, risk, astart, aend, p, alpha=0.05, power=0.8, eage=NULL,
          agegrp=NULL)
```

Arguments

eexpo	the design value of the exposure related relative incidence.
risk	a positive number showing the duration the risk period. It should not be greater than the duration of the shortest age group, if age groups are specified.
astart	age at start of the observation period. It is an integer greater or equal to 0, that is the same start of observation for all cases.
aend	age at end of an observation period. It is an integer greater than 0, that is the same end of observation for all cases.
p	a vector/scalar of proportions of exposed in each age group, the default is p=1 where there are no age effects and all cases are exposed.
alpha	level of significance e.g 0.05.
power	the power required to detect the relative incidence specified in eexpo, e.g 0.8.
eage	age related relative incidence parameters, the reference group is the first one.
agegrp	a vector of cut points of the age groups where each value represents the start of an age category. The first element in the vector is the start of the second age group. The first age group starts at the minimum of astart, the start of observation period. The default is NULL (i.e no age effects included).

Value

a sample size is produced.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Musonda, P., Farrington, C. P., and Whitaker, H. (2006). Sample sizes for self-controlled case series studies. Statistics in Medicine 25, 2618–2631.

Farrington P., Whitaker H., and Ghebremichael-Weldeselassie Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

Examples

```
# Sample size for exposure RI = 2.5 with 21 days risk period,
# all cases exposed. The level of significance is 0.05
# with 80% power. The sample size of when p=1 is:

   ss1 <- samplesize(eexpo=2.5, risk=21, astart=366, aend=730,
                p=1, alpha=0.05, power=0.8)

   ss1


# the sample size of events from cases exposed or not when 75% of the
# population are exposed
```

```
   ss2 <- samplesize(eexpo=2.5, risk=21, astart=366, aend=730,
                p=0.75, alpha=0.05, power=0.8)


   ss2

# Sample size when age effect is included and the proportions of the
# target exposed population which are exposed in each age group
# are p=c(0.50,0.35,0.1,0.05):

   ss3 <- samplesize(eexpo=2.5, risk=21, astart=366, aend=730,
                p=c(0.50,0.35,0.1,0.05), alpha=0.05, power=0.8,
                eage=c(1.2,1.6,2.0), agegrp=c(457,548,639))

   ss3

# Suppose that the sample is from the entire population with 75% exposed,
# then p=0.75*c(0.50,0.35,0.1,0.05)

  ss4 <- samplesize(eexpo=2.5, risk=21, astart=366, aend=730,
                p=0.75*c(0.50,0.35,0.1,0.05), alpha=0.05, power=0.8,
                eage=c(1.2,1.6,2.0), agegrp=c(457,548,639))
  ss4
```

`semisccs` *Semiparametric self-controlled case series method*

Description

The function fits the semiparametric self-controlled case series method where the age effect is left unspecified, as published in Farrington and Whitaker (2006).

Usage

```
semisccs(formula, indiv, astart, aend, aevent, adrug, aedrug, expogrp = list(),
            washout = list(), sameexpopar = list(), dataformat="stack", data)
```

Arguments

`formula`	model formula. The dependent variable should always be "event" e.g. event ~ itp. There is no need to specify age in the model formula.
`indiv`	a vector of individual identifiers of cases.
`astart`	a vector of ages at which the observation periods start.
`aend`	a vector of ages at end of observation periods.
`aevent`	a vector of ages at event, an individual can experience multiple events.

adrug	a list of vectors of ages at start of exposures or a list of matrices if the exposures have multiple episodes (dataformat multi). Multiple exposures of the same type can be recorded as multiple rows (dataformat stack). One list item per exposure type.
aedrug	a list of vectors of ages at which exposure-related risk ends or a list of matrices if there are multiple episodes (repeat exposures in different columns) of the same exposure type. The dimension of each item of aedrug has to be equal to that of adrug, that is aedrug should be given for each exposure in adrug.
expogrp	list of vectors of days to the start of exposure-related risk, counted from adrug. E.g if the risk period is [adrug+c,aedrug], use expogrp = c. For multiple exposure types expogrp is a list of length as list adrug. The DEFAULT is a list of zeros where the exposure-related risk periods are [adrug, aedrug].
washout	list of vectors with days to start of washout periods counted from aedrug, the number of vectors in the list is equal to the number of exposure types or the length of adrug. The default is NULL, no washout periods. The order of the list items corresponds to the order of exposures in adrug.
sameexpopar	a vector of logical values. If TRUE (the default) no dose effect is assumed: the same exposure parameters are used for multiple doses of the same exposure type, presented in dataformat 'multi'. If FALSE different relative incidences are estimated for different doses of the same exposure type. The length of the vector is equal to the length list adrug.
dataformat	the way the input data are assembled. It accepts "multi" or "stack" (the default), where "multi" refers to a data assembled with one row representing one event and "stack" refers to a data frame where repeated exposures of the same type are stack in one column. In the "multi" dataformat different episodes of the same type are recorded as separate columns in the dataframe.
data	a data frame containing the input data. The data should be in 'stack' or 'multi' (see dataformat).

Details

In the standard SCCS method both age and exposure effects are modelled using step functions. However, mis-specification of age groups in the standard SCCS may lead to bias in the exposure related relative incidence estimates. In the semiparametric SCCS no age groups are pre-specified. A parameter for each day an event occurred is fitted, which means that this method is only suitable for small to medium sized data sets. An alternative for large data sets is provided by smoothagesccs.

Value

The function returns age and exposure related relative incidence estimates along with 95% confidence limits.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington, C. P., Whitaker, H. J. (2006). Semiparametric analysis of case series data. Applied Statistics, 55(5): 553–594.

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

standardsccs, smoothagesccs, smoothexposccs

Examples

```
# Example 1
# Semiparametric model for the ITP and MMR vaccine data

itp.mod1 <- semisccs(event~mmr, indiv=case, astart=sta,
                     aend=end, aevent=itp, adrug=mmr, aedrug=mmr+42,
                     expogrp=c(0,15,29), data=itpdat)

itp.mod1


# Example 2
# Data on itp and mmr vaccine
# Sex and mmr interaction included

itp.mod2 <- semisccs(event~factor(sex)*mmr, indiv=case,
                     astart=sta, aend=end, aevent=itp, adrug=mmr,
                     aedrug=mmr+42, expogrp=c(0,15,29), data=itpdat)


itp.mod2
```

---

siddat *Data on hexavalent vaccine and sudden infant death syndrome (SIDS)*

Description

These simulated data comprise ages in days at hexavalent vaccination and SIDS in 300 cases.

Usage

```
siddat
```

Format

A data frame containing 300 rows and 5 columns. The column names are 'case' (individual identifier), 'sta' (age on first day of the observation period), 'end' (age on last day of the nominal observation period), 'sids' (age at SIDS), 'hex' (age at first hexavalent vaccination), 'hexd2' (age at second dose), 'hexd3' (age at third dose).

References

Kuhnert R., Hecker, H., Poethko-Muller, C., Schlaud, M., Vennemann, M., Whitaker, H.J., and Farrington C. P. (2011). A modified self-controlled case series method to examine association between multidose vaccinations and death. Statistics in Medicine 30, 666-677.

---

simulatesccsdata *Simulation of SCCS data*

Description

This function creates a simulated SCCS data set with given design parameters, and can be used to generate cases with observation and risk periods of different durations, multiple risk periods, repeated exposures, and washout periods.

Usage

```
simulatesccsdata(nindivs, astart, aend, adrug, aedrug, expogrp=c(0), eexpo,
                 washout=NULL, ewashout=NULL, agegrp=NULL, eage=NULL)
```

Arguments

nindivs	a positive integer: number of cases to be generated (1 event per case).
astart	age at start of an observation period. It is a single number if the same start of observation for all cases is required or a vector of length equal to nindivs to allow different starts of observation periods for different cases.
aend	age at end of the observation period. A single number for the same end of observation periods for all cases or a vector to allow for different end of observation periods.
adrug	a vector (of length nindivs) of ages at which exposure starts or a matrix if there are multiple exposures.
aedrug	a vector of ages at which exposure-related risk ends or a matrix if there are multiple exposures. The number of columns of aedrug is equal to the number of columns of adrug, that is aedrug should be given for each column in adrug.
expogrp	a vectors of days to the start of exposure-related risk, counted from adrug. E.g if the risk period is [adrug+c,aedrug], use expogrp = c.
eexpo	a vector of exposure-related relative incidences.
washout	a vector of days to start of washout periods counted from aedrug. The default is NULL, no washout periods.

ewashout	a vector of true relative incidence values associated with washout periods; it defaults to NULL when washout=NULL.
agegrp	cut points of age groups, defaults to NULL (i.e no age effect included). These are given as the day of an age category starts, the first age category starts at the minimum of astart.
eage	a vector of age-related relative incidences. The default is NULL where there is no age effect i.e agegrp = NULL. If age-specific relative incidences are from a continuous function eage is a vector of relative incidences at each age and agegrp=NULL.

Details

The true relative incidences related to age and exposure could be generated from discrete or continuous distributions.

Value

A data frame with columns "indiv" = individual identifier, "astart" = age on the day observation period starts, "adrug" = age on the day exposure starts, "aedrug" = age at the end of exposure related risk period, "aend" = age at the end of observation period, and "aevent" = age on the day of outcome event.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

Examples

```
# Simulate data where all the cases have same start and end of
# observation periods and no age effect

set.seed(4321)

arisk <- round(runif(110,366,730)) # ages at start of exposure

simdata <- simulatesccsdata(nindivs=110, astart=366, aend=730,
                            adrug=arisk, aedrug=arisk+20, eexpo=2.5)


simdata
```

`smoothagesccs` *Spline-based semiparametric SCCS, smooth age*

Description

Fits a semiparametric SCCS model with smooth age effect, where the age related relative incidence function is represented by spline function; that is, linear combinations of M-splines. The exposure related relative incidence function is represented by step functions. One exposure group can be included.

Usage

```
smoothagesccs(indiv, astart, aend, aevent, adrug, aedrug, expogrp = 0,
              washout = NULL, kn=12, sp = NULL, data)
```

Arguments

`indiv`	a vector of individual identifiers of cases.
`astart`	a vector of ages at which observation periods start.
`aend`	a vector of ages at end of observation periods.
`aevent`	a vector of ages at event, an individual can experience multiple events.
`adrug`	a vector of ages at which exposure starts, only a single exposure type can be included.
`aedrug`	a vector of ages at which the exposure-related risk periods end.
`expogrp`	a vector of days to the start of exposure-related risk, counted from `adrug`. E.g if the risk period is [`adrug`+c,`aedrug`], use expogrp = c. To define multiple risk windows, `expogrp` is a vector of days on start of risk periods counted from `adrug`. The DEFAULT is zero where the exposure-related risk periods are [`adrug`, `aedrug`].
`washout`	a vector of days to start of washout periods counted from `aedrug`. The default is NULL, no washout periods.
`kn`	an integer >=5 representing the number of interior knots used to define the M-spline basis functions which are related to the age specific relative incidence function, usually between 8 and 12 knots is sufficient. It defaults to 12 knots.
`sp`	smoothing parameter value. It defaults to "auto" where the smoothing paramter is obtained automatically using a cross-validation method. The value of "sp" must be a number greater or equal to 0.
`data`	a data frame containing the input data. The data are assembled one line per event.

Details

The standard SCCS represents the age and exposure effects by piecewise constant step functions, however mis-specification of age group cut points might lead to biased estimates of the exposure related relative incidences. The semiparametric SCCS model, `semisccs`, has numerical challenges when the number of cases is large. This splined-based semiparametric SCCS model with smooth age effect avoids these limitations of the standard and semiparametric SCCS models. The smoothing parameter for the age-related relative incidence function is chosen by an approximate cross-validation method. The method is outlined in Ghebremichael-Weldeselassie et al (2014).

Value

Relative incidence estimates along with their 95% confidence limits.

`coef`	log of the exposure related relative incidence estimates.
`se`	standard errors of the log of exposure related relative incidence estimates.
`age`	age related relative incidences at each day between the minimum age at start of observation and maximum age at end of observation periods.
`ageaxis`	sequence of ages between the minimum age at start of observations and maximum age at end of observation periods corresponding to the age related relative incidences.
`smoothingpara`	smoothing parameter chosen by maximizing an approximate cross-validation score or given as an argument in the function
`cv`	cross-validation score

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Ghebremichael-Weldeselassie, Y., Whitaker, H. J., Farrington, C. P. (2015). Self-controlled case series method with smooth age effect. Statistics in Medicine, 33(4), 639-649.

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

`smoothexposccs`

Examples

```
# Fit the SCCS model with smooth age effect to the itp data and plot age effect.

itp.mod <- smoothagesccs(indiv=case, astart=sta,aend=end, aevent=itp,
                         adrug=mmr, aedrug=mmr+42, expogrp=c(0,15,29), sp=2800,
                         data=itpdat)

itp.mod
```

```
plot(itp.mod)
```

`smoothexposccs` *Spline-based semiparametric SCCS, smooth exposure*

Description

Fits a spline-based SCCS model where the exposure-related relative incidence function is represented by a spline function, that is a linear combination of M-splines, and the age effects are represented by a piecewise constant function.

Usage

```
smoothexposccs(indiv, astart, aend, aevent, adrug, aedrug, agegrp, kn=12,
               sp = NULL, data)
```

Arguments

`indiv`	a vector of individual identifiers of cases.
`astart`	a vector of ages at which the observation periods start.
`aend`	a vector of ages at end of observation periods.
`aevent`	a vector of ages at event (outcome of interest), an individual can experience multiple events.
`adrug`	a vector of ages at which exposure related risk period starts.
`aedrug`	a vector of ages at which exposure related risk period ends.
`agegrp`	a vector of cut points for the age groups where each value represents the start of an age catagory. The first element in the vector is the start of the second age group. The first age group starts at the minimum of `astart`, the start of the observation period.
`kn`	number of interior knots >=5 used to define the M-spline basis functions, usually between 8 and 12 knots is sufficient. The default is 12.
`sp`	smoothing parameter value. It defaults to "auto" where the smoothing paramter is obtained automatically using a cross validation method. The value of "sp" must be a number greater or equal to 0.
`data`	a data frame containing the input data.

Details

The `standardsccs`, `semisccs` and `smoothagesccs` use piecewise constant step functions to model the exposure effect. However mis-specification of exposure group cut points might result in biased estimates. This method represents exposure related relative incidence function by a spline function.

Value

Relative incidence estimates along with their 95% confidence limits. Varaince-covariance matrix can also be obtained.

estimates	exposure related relative incidence estimates at each point of time since start of exposure until the maximum duration of exposure.
lci	lower confidence limits of the exposure-related relative incidence estimates.
uci	upper confidence limits of the exposure-related relative incidence estimates.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Ghebremichael-Weldeselassie, Y., Whitaker, H. J., Farrington, C. P. (2015). Flexible modelling of vaccine effects in self-controlled case series models 25, 1768–1797.

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series Studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

smoothagesccs, nonparasccs

Examples

```
library(SCCS)

# Fit smooth exposure SCCS to MMR vaccine and itp

 itp.mod1 <- smoothexposccs(sp=10, indiv=case, astart=sta, aend=end,
                            aevent=itp, adrug=mmr, aedrug=mmr+42,
                            agegrp=c(427, 488, 549, 610, 671), data=itpdat)

 itp.mod1

 plot(itp.mod1)
```

---

standardsccs *The standard SCCS method*

Description

Fits the standard SCCS model where age and exposure effects are represented by piecewise constant functions.

Usage

```
standardsccs(formula, indiv, astart, aend, aevent, adrug, aedrug,
            expogrp = list(), washout = list(), sameexpopar = list(),
            agegrp = NULL, seasongrp=NULL, dob=NULL, dataformat="stack",
            data)
```

Arguments

formula	model formula. The dependent variable should always be "event" e.g. event ~ itp. If age and/or season effects are included, they should always be included as 'age' and 'season', e.g. event ~ itp + age.
indiv	a vector of individual identifiers of cases.
astart	a vector of ages at which the observation periods start.
aend	a vector of ages at end of observation periods.
aevent	a vector of ages at event; an individual can experience multiple events
adrug	a list of vectors of ages at start of exposures or a list of matrices if the exposures have multiple episodes (dataformat multi). Multiple exposures of the same type can be recorded as multiple rows (dataformat stack). One list item per exposure type.
aedrug	a list of vectors of ages at which exposure-related risk ends or a list of matrices if there are multiple episodes (repeat exposures in different columns) of the same exposure type. The dimension of each item of aedrug has to be equal to that of adrug, that is aedrug should be given for each exposure in adrug.
expogrp	list of vectors of days to the start of exposure-related risk, counted from adrug. E.g if the risk period is [adrug+c,aedrug], use expogrp = c or expogrp = list(c). For multiple exposure types expogrp is a list of the same length as list adrug. The DEFAULT is a list of zeros where the exposure-related risk periods are [adrug, aedrug].
washout	list of vectors with days to start of washout periods counted from aedrug, the number of vectors in the list is equal to the number of exposures or the length of list adrug. The default is NULL, no washout periods. The order of the list corresponds to the order of exposures in adrug.
sameexpopar	a vector of logical values. If TRUE (the default) no dose effect is assumed: the same exposure parameters are used for multiple doses/episodes of the same exposure type, presented in dataformat 'multi'. If FALSE different relative incidences are estimated for different doses of the same exposure. The length of the vector is equal to the length of list adrug. The order of the elements of the vector corrosponds to the order of exposures in list adrug.
agegrp	a vector of cut points of the age groups where each value represents the start of an age category. The first element in the vector is the start of the second age group. The first age group starts at the minimum of astart, the start of the observation period. The default is NULL (i.e no age effects included).
seasongrp	a vector of cut points for seasonal effects. The values should be given in ddmm format, representing the first days of each season group. The seasonal effect is a factor, the reference level being the time interval starting at the earliest date in seasongrp. The default is NULL where no seasonal effects are included.

dob	a vector of birth dates of the cases, in ddmmyyyy format. They are used if seasonal effects are included in the model. The default dob is NULL. Required if seasongrp is not NULL.
dataformat	the way the input data are assembled. It accepts "multi" or "stack" (the default), where "multi" refers to a data assembled with one row representing one event and "stack" refers to a data frame where repeated exposures of the same type are stack in one column. In the "multi" dataformat different episodes of the same exposure type are recorded as separate columns in the dataframe.
data	a data frame containing the input data. The data should be in 'stack' or 'multi' (see dataformat).

Details

In the standard SCCS model, originally described in Farrington (1995), age and exposure effects are represented by step functions. Suppose that individual i has n_i events, n_{ijk} occurring in age group j and exposure group k and that the s^{th} event falls within age group j_s and exposure group k_s. The SCCS likelihood contribution for individual i is

$$l_i = \frac{\prod_{s=1}^{n_i} \exp(\alpha_{j_s} + \beta_{k_s})}{\left(\sum_{j=1}^{J} \sum_{k=1}^{K} \exp(\alpha_j + \beta_k) e_{ijk} \right)^{n_i}}.$$

This is a multinomial likelihood with index n_i, responses n_{ijk} and probabilities

$$p_{iuv} = \frac{\exp(\alpha_u + \beta_v) e_{iuv}}{\sum_{j=1}^{J} \sum_{k=1}^{K} \exp(\alpha_j + \beta_k) e_{ijk}}.$$

The standard SCCS likelihood is equivalent to a product multinomial likelihood.

The SCCS likelihood is equivalent to that of the conditional logistic model for $1 : M$ matched case-control studies: see Breslow and Day (1980), Chapter 7. This means that SCCS models can be fit using conditional logistic regression, with a factor for each individual event.

The SCCS R package maximises the likelihood using clogit. Each event is assigned an identifier (indivL), and direct estimation is avoided using the option strata(indivL).

Value

Relative incidence estimates along with their 95% confidence limits.

Author(s)

Yonas Ghebremichael-Weldeselassie, Heather Whitaker, Paddy Farrington.

References

Farrington, C.P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. Biometrics 51, 228–235.

Breslow, N. E. and Day, N. E. (1980). Statistical Methods in Cancer Research, volume I: The analysis of case-control studies. IARC Publications No.32.

Farrington, P., Whitaker, H., and Ghebremichael-Weldeselassie, Y. (2018). Self-controlled Case Series studies: A modelling Guide with R. Boca Raton: Chapman & Hall/CRC Press.

See Also

semisccs

Examples

```
# Single exposure-related risk period with no age effect

  itp.mod1 <- standardsccs(event~mmr, indiv=case, astart=sta, aend=end,
                       aevent=itp, adrug=mmr, aedrug=mmr+42, data=itpdat)

  itp.mod1

# Single exposure-related risk period and age effect included

  itp.mod2 <- standardsccs(event~mmr+age, indiv=case, astart=sta, aend=end,
                       aevent=itp, adrug=mmr, aedrug=mmr+42,
                       agegrp=c(427,488,549,610,671), data=itpdat)

  itp.mod2

# Multiple risk periods and age effect included

  itp.mod3 <- standardsccs(event~mmr+age, indiv=case, astart=sta, aend=end,
                       aevent=itp, adrug=mmr, aedrug=mmr+42, expogrp=c(0,15,29),
                       agegrp=c(427,488,549,610,671), data=itpdat)

  itp.mod3

# Multiple risk periods, washout periods and age effects

   ageq <- floor(quantile(hipdat$frac, seq(0.05,0.95,0.05),
                        names=FALSE)) # Age group
                                    # cut points

   hip.mod1 <- standardsccs(event~ad+age, indiv=case, astart=sta, aend=end,
                          aevent=frac, adrug=ad, aedrug=endad, expogrp=c(0,15,43),
                          washout=c(1,92,182), agegrp=ageq, data=hipdat)


# Multiple/repeat exposures of the same exposure type, dataformat="stack"

   ageq <- floor(quantile(gidat$bleed[duplicated(gidat$case)==0],
                seq(0.025,0.975,0.025), names=FALSE))

   gi.mod1 <- standardsccs(event~ns+relevel(age,ref=21), indiv=case, astart=sta,
                         aend=end, aevent=bleed, adrug=ns, aedrug=endns,
                         agegrp=ageq, dataformat="stack", data=gidat)

   gi.mod1
```

```
#  Multiple doses of a vaccine each with different parameter estimates (sameexpopar=F)


   ageg <- c(57,85,113,141,169,197,225,253,281,309,337) # age group cut points

   dtp.mod2 <- standardsccs(event~dtp+age, indiv=case, astart=sta, aend=end,
                            aevent=conv, adrug=cbind(dtp,dtpd2, dtpd3),
                            aedrug=cbind(dtp+14,dtpd2+14,dtpd3+14),
                            expogrp=c(0,4,8),agegrp=ageg, dataformat="multi",
                            sameexpopar=FALSE, data=dtpdat)

   dtp.mod2

# Multiple exposure types

  ageg <- seq(387,707,20) # Age group cut points
 con.mod <- standardsccs(event~hib+mmr+age, indiv=case, astart=sta, aend=end,
                         aevent=conv, adrug=cbind(hib,mmr), aedrug=cbind(hib+14,mmr+14),
                          expogrp=list(c(0,8), c(0,8)), agegrp=ageg, data=condat)

  con.mod


# Multiple doses/episodes of several exposure types, the doses of each exposure type
# have same paramter


   ageg <- c(57,85,113,141,169,197,225,253,281,309,337)   # age group cut points

   hib.mod1 <- standardsccs(event~dtp+hib+age, indiv=case, astart=sta,
                            aend=end, aevent=conv,
                           adrug=list(cbind(dtp, dtpd2,dtpd3),
                                      cbind(hib,hibd2,hibd3)),
                            aedrug=list(cbind(dtp+14,dtpd2+14,dtpd3+14),
                                        cbind(hib+14,hibd2+14,hibd3+14)),
                            expogrp=list(c(0,4,8),c(0,8)),agegrp=ageg,
                            dataformat="multi", data=hibdat)

   hib.mod1


# Multiple doses/episodes of several exposure types, the doses of "dtp"
# different parameters and the doses of the second exposure hib have
# same paramters


   ageg <- c(57,85,113,141,169,197,225,253,281,309,337)
                  # age group cut points

   hib.mod2 <- standardsccs(event~dtp+hib+age,
                                  indiv=case, astart=sta, aend=end,
                       aevent=conv, adrug=list(cbind(dtp,dtpd2,dtpd3),
```

```
                                cbind(hib,hibd2,hibd3)),
                        aedrug=list(cbind(dtp+3,dtpd2+3,dtpd3+3),
                                cbind(hib+7,hibd2+7,hibd3+7)),
                        sameexpopar=c(FALSE,TRUE), agegrp=ageg,
                        dataformat="multi", data=hibdat)
    hib.mod2


# Season included in a model

    month <- c(0101,0102,0103,0104,0105,0106,0107,0108,0109,0110,0111,0112)
    # season cutpoints

    int.mod <- standardsccs(event~opv+age+season, indiv=case, astart=sta,
                            aend=end, aevent=intus, adrug=cbind(opv,opvd2),
                            aedrug=cbind(opv+42,opvd2+42), expogrp=c(0,15,29),
                            agegrp=seq(30,330,30), seasongrp=month,dob=dob,
                            dataformat="multi", data=intdat)
    int.mod
```

Index

∗ **Event dependent exposures**
eventdepenexp, 8
∗ **Parametric**
eventdepenobs, 10
quantsccs, 26
standardsccs, 38
∗ **Sample size**
samplesize, 28
∗ **Semi-parametric SCCS**
smoothexposccs, 37
∗ **Semi-parametric**
semisccs, 30
smoothagesccs, 35
∗ **Splines**
integrateIspline, 19
nonparasccs, 22
smoothagesccs, 35
smoothexposccs, 37
∗ **datasets**
addat, 3
adidat, 4
amdat, 4
apdat, 5
autdat, 6
bpdat, 6
bupdat, 7
condat, 7
dtpdat, 8
febdat, 13
gbsdat, 16
gidat, 17
hibdat, 18
hipdat, 18
intdat, 19
itpdat, 20
midat, 22
nrtdat, 24
opvdat, 25
pmdat, 26
rotdat, 27
rsvdat, 28
siddat, 32
∗ **likelihodd ratio test**
lrtsccs, 21
∗ **package**
SCCS-package, 2

addat, 3
adidat, 4
amdat, 4
apdat, 5
autdat, 6

bpdat, 6
bupdat, 7

condat, 7

dtpdat, 8

eventdepenexp, 8
eventdepenobs, 10

febdat, 13
formatdata, 14

gbsdat, 16
gidat, 17

hibdat, 18
hipdat, 18

intdat, 19
integrateIspline, 19
itpdat, 20

lrtsccs, 21

midat, 22

nonparasccs, 22, *38*

nrtdat, 24

opvdat, 25

pmdat, 26

quantsccs, 26

rotdat, 27
rsvdat, 28

samplesize, 28
SCCS *(*SCCS-package*)*, 2
SCCS-package, 2
semisccs, *10*, *27*, 30, *37*, *41*
siddat, 32
simulatesccsdata, 33
smoothagesccs, *24*, *31*, *32*, 35, *37*, *38*
smoothexposccs, *24*, *32*, *36*, 37
standardsccs, *10*, *21*, *32*, *37*, 38