Package: serofoi (via r-universe)

November 5, 2024

Title Estimates the Force-of-Infection of a Given Pathogen from Population Based Seroprevalence Studies

Version 0.1.0

Description Estimate time-varying Force-of-Infection of a given pathogen from population based seroprevalence studies using a bayesian framework.

License MIT + file LICENSE

URL `https://epiverse-trace.github.io/serofoi/`, `https://github.com/epiverse-trace/serofoi`

BugReports `https://github.com/epiverse-trace/serofoi/issues`

Depends R (>= 3.5.0)

Imports bayesplot, checkmate, cowplot, dplyr, ggplot2, Hmisc, loo, methods, purrr, Rcpp (>= 0.12.0), rstan (>= 2.18.1), rstantools (>= 2.3.1)

Suggests knitr, rmarkdown, spelling, testthat (>= 3.0.0), vdiffr

LinkingTo BH (>= 1.66.0), Rcpp (>= 0.12.0), RcppEigen (>= 0.3.3.3.0), RcppParallel (>= 5.0.1), rstan (>= 2.18.1), StanHeaders (>= 2.18.0)

VignetteBuilder knitr

Additional_repositories `https://mc-stan.org/r-packages/`

Biarch true

Config/Needs/website epiverse-trace/epiversetheme

Config/testthat/edition 3

Encoding UTF-8

Language en-GB

LazyData true

NeedsCompilation yes

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

SystemRequirements GNU make

Config/pak/sysreqs make libicu-dev

Repository https://epiverse-trace.r-universe.dev

RemoteUrl https://github.com/epiverse-trace/serofoi

RemoteRef HEAD

RemoteSha a608a0ddd2221ed129aa583a6d67a91eb8eb2634

Contents

`serofoi-package` *The 'serofoi' package.*

Description

A DESCRIPTION OF THE PACKAGE

Author(s)

Maintainer: Zulma M. Cucunubá `<zulma.cucunuba@javeriana.edu.co>` (ORCID)

Authors:

- Nicolás T. Domínguez `<ex-ntorres@javeriana.edu.co>` (ORCID)
- Ben Lambert
- Pierre Nouvellet

References

Stan Development Team (NA). RStan: the R interface to Stan. R package version 2.26.22. https://mc-stan.org

See Also

Useful links:

- `https://trace-lac.github.io/serofoi/`

`chagas2012` *Seroprevalence data on serofoi*

Description

Data from a serological surveys

Usage

```
chagas2012
```

Format

An object of class `"cross"`; see `qtl::read.cross()`.

Examples

```
chagas2012
```

`chik2015` *Seroprevalence data on serofoi*

Description

Data from a serological surveys

Usage

```
chik2015
```

Format

An object of class "`cross`"; see `qtl::read.cross()`.

Examples

```
chik2015
```

`extract_seromodel_summary` *Function to extract a summary of the specified serological model object*

Description

This function extracts a summary corresponding to a serological model object containing information about the original serological survey data used to fit the model, such as the year when the survey took place, the type of test taken and the corresponding antibody, as well as information about the convergence of the model, like the expected log pointwise predictive density `elpd` and its corresponding standard deviation.

Usage

```
extract_seromodel_summary(seromodel_object, serodata)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place

`age_mean_f`	Floor value of the average between age_min and age_max
`sample_size`	The size of the sample
`birth_year`	The year in which the individuals of each age group were born

The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.

Value

`model_summary`. Object with a summary of `seromodel_object` containing the following:

`foi_model`	Name of the selected model.
`data_set`	Seroprevalence survey label
`country`	Name of the country were the survey was conducted in.
`year`	Year in which the survey was conducted.
`test`	Type of test of the survey.
`antibody`	Antibody
`n_sample`	Total number of samples in the survey.
`n_agec`	Number of age groups considered.
`n_iter`	Number of iterations by chain including warmup.
`elpd`	elpd
`se`	se
`converged`	convergence

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant"
)
extract_seromodel_summary(seromodel_object,
  serodata = serodata
)
```

`fit_seromodel` *Fit selected model to the specified seroprevalence survey data*

Description

This function fits the specified model `foi_model` to the serological survey data `serodata` by means of sampling. The function determines whether the corresponding stan model object needs to be compiled by rstan.

Usage

```
fit_seromodel(
  serodata,
  foi_model = c("constant", "tv_normal_log", "tv_normal"),
  foi_location = 0,
  foi_scale = 1,
  chunks = NULL,
  chunk_size = 1,
  iter = 1000,
  adapt_delta = 0.9,
  max_treedepth = 10,
  seed = 12345,
  ...
)
```

Arguments

`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`foi_model`	Name of the selected model. Current version provides three options: `"constant"` Runs a constant model `"tv_normal"` Runs a normal model `"tv_normal_log"` Runs a normal logarithmic model
`foi_location`	Location parameter of the force-of-infection distribution of the selected model. Depending on `foi_model`, the meaning may vary.
`foi_scale`	Scale parameter of the force-of-infection distribution of the selected model. Depending on `foi_model`, the meaning may vary.
`chunks`	Numeric list specifying the chunk structure of the time interval from the birth year of the oldest age cohort `min(serodata$age_mean_f)` to the time when the serosurvey was conducted `t_sur`. If `NULL`, the time interval is divided in chunks of size `chunk_size`.
`chunk_size`	Size of the chunks to be used in case that the chunk structure `chunks` is not specified in fit_seromodel. Default is 1, meaning that one force of infection value is to be estimated for every year in the time interval spanned by the serosurvey. If the length of the time interval is not exactly divisible by `chunk_size`, the remainder years are included in the last chunk.
`iter`	Number of interactions for each chain including the warmup. `iter` in sampling.

`adapt_delta`	Real number between 0 and 1 that represents the target average acceptance probability. Increasing the value of `adapt_delta` will result in a smaller step size and fewer divergences. For further details refer to the `control` parameter in sampling or here.
`max_treedepth`	Maximum tree depth for the binary tree used in the NUTS stan sampler. For further details refer to the `control` parameter in sampling.
`seed`	For further details refer to the `seed` parameter in sampling.
`...`	Additional parameters for sampling.

Value

`seromodel_object`. `stanfit` object returned by the function sampling

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_fit <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant"
)
```

`generate_sim_data` *Generate simulated serosurvey according to the specified FoI*

Description

Generate simulated serosurvey according to the specified FoI

Usage

```
generate_sim_data(
  sim_data,
  foi,
  sample_size_by_age,
  mu = 0,
  model_type = "time-varying",
  survey_label = "sim_data",
  seed = 1234
)
```

Arguments

`sim_data`	A dataframe object containing the following columns: `age` Age group markers `tsur` Year of the survey
`foi`	Numeric atomic vector corresponding to the desired time-varying or age-varying Force-of-Infection to simulate from
`sample_size_by_age`	Integer indicating the sample size by age group. This corresponds to the number of trials `size` in rbinom.
`mu`	Seroreversion rate
`model_type`	String specifying the type of model to be used. Current valid options are 'time-varying' and 'age-varying'
`survey_label`	Label for the resulting simulated serosurvey.
`seed`	Seed for random number generation.

Value

Dataframe containing the simulated serosurvey.

Examples

```
n_years <- 50
sim_data <- data.frame(
  age = seq(1,n_years),
  tsur = 2050
)
foi <- rep(0.02, n_years)
sample_size_by_age <- as.integer(runif(n = n_years, 5, 10))
sim_data <- generate_sim_data(
  sim_data = sim_data,
  foi = foi,
  sample_size_by_age = sample_size_by_age,
  survey_label = "sim_constant_foi"
)
```

`get_age_group` *Construct age-group variable from age column*

Description

Simplified version of get_age_group. This function splits an age interval from age_min to age_max by steps of length `step`. `age_min` and `age_max` are calculated from `age`. In cases that `age_max%%(step+1)!=0`, the last age interval is truncated and will have a different length than the others.

Usage

```
get_age_group(age, step)
```

Arguments

`age`	vector containing age information
`step`	step used to split the age interval

Value

age_group factor variable grouping age by the age intervals specified by `min(age)`, `max(age)`.

`get_chunk_structure` *Generate list containing the chunk structure to be used in the retrospective estimation of the force of infection.*

Description

This function generates a numeric list specifying the chunk structure of the time interval spanning from the year of birth of the oldest age cohort up to the time when the serosurvey was conducted.

Usage

```
get_chunk_structure(serodata, chunk_size)
```

Arguments

`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`chunk_size`	Size of the chunks to be used in case that the chunk structure `chunks` is not specified in fit_seromodel. Default is 1, meaning that one force of infection value is to be estimated for every year in the time interval spanned by the serosurvey. If the length of the time interval is not exactly divisible by `chunk_size`, the remainder years are included in the last chunk.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(serodata = chagas2012, alpha = 0.05)
cohort_ages <- get_cohort_ages(serodata = serodata)
```

`get_cohort_ages` *Generate data frame containing the age of each cohort corresponding to each birth year excluding the year of the survey.*

Description

This function generates a data frame containing the age of each cohort corresponding to each `birth_year` excluding the year of the survey, for which the cohort age is still 0. specified serological survey data `serodata` excluding the year of the survey.

Usage

```
get_cohort_ages(serodata)
```

Arguments

`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns:

- `total` Number of samples for each age group
- `counts` Number of positive samples for each age group
- `tsur` Year in which the survey took place
- `age_mean_f` Floor value of the average between age_min and age_max
- `sample_size` The size of the sample
- `birth_year` The year in which the individuals of each age group were born

The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.

Value

`cohort_ages`. A data frame containing the age of each cohort corresponding to each birth year

Examples

```
data(chagas2012)
serodata <- prepare_serodata(serodata = chagas2012, alpha = 0.05)
cohort_ages <- get_cohort_ages(serodata = serodata)
```

`get_foi_central_estimates` *Extract central estimates for the fitted forced FoI*

Description

Extract central estimates for the fitted forced FoI

Usage

```
get_foi_central_estimates(seromodel_object, cohort_ages)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`cohort_ages`	A data frame containing the age of each cohort corresponding to each birth year.

Value

`foi_central_estimates`. Central estimates for the fitted forced FoI

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant"
)
cohort_ages <- get_cohort_ages(serodata = serodata)
foi_central_estimates <- get_foi_central_estimates(
  seromodel_object = seromodel_object,
  cohort_ages = cohort_ages
)
```

`get_prev_expanded` *Generate data frame containing the confidence interval based on a force-of-infection fitting*

Description

This function computes the corresponding binomial confidence intervals for the obtained prevalence based on a fitting of the force-of-infection `foi` for plotting an analysis purposes.

Usage

```
get_prev_expanded(foi, serodata, alpha = 0.05, bin_data = FALSE, bin_step = 5)
```

Arguments

`foi`	Object containing the information of the force-of-infection. It is obtained from `rstan::extract(seromodel_object$seromodel, "foi", inc_warmup = FALSE)[[1]]`.
`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`alpha`	Probability threshold for statistical significance used for both the binomial confidence interval, and the lower and upper quantiles of the estimated prevalence.
`bin_data`	If `TRUE`, `serodata` is binned by means of `prepare_bin_data`. Otherwise, age groups are kept as originally input.
`bin_step`	Integer specifying the age groups bin size to be used when `bin_data` is set to `TRUE`.

Value

`prev_final`. The expanded prevalence data. This is used for plotting purposes in the `visualization` module.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant"
)
foi <- rstan::extract(seromodel_object, "foi")[[1]]
get_prev_expanded(foi, serodata)
```

`get_sim_n_seropositive` *Generate sample of counts of seropositive individuals by sampling from a binomial distribution*

Description

Generate sample of counts of seropositive individuals by sampling from a binomial distribution

Usage

```
get_sim_n_seropositive(
  sim_data,
  foi,
  sample_size_by_age,
  mu = 0,
  model_type = "time-varying",
  seed = 1234
)
```

Arguments

`sim_data`	A dataframe object containing the following columns: `age` Age group markers `tsur` Year of the survey
`foi`	Numeric atomic vector corresponding to the desired time-varying or age-varying Force-of-Infection to simulate from
`sample_size_by_age`	Integer indicating the sample size by age group. This corresponds to the number of trials `size` in rbinom.
`mu`	Seroreversion rate
`model_type`	String specifying the type of model to be used. Current valid options are 'time-varying' and 'age-varying'
`seed`	Seed for random number generation.

Value

A dataframe containing the following columns:

`age` Age by the time of the survey

`n_seropositive` Number of positive cases sampled according to the provided FoI

simulated list of counts following a binomial distribution in accordance with a given force of infection and age group sizes.

Examples

```
n_years <- 50
sim_data <- data.frame(
  age = seq(1,n_years),
  tsur = 2050
)
foi <- rep(0.02, n_years)
sample_size_by_age <- as.integer(runif(n = n_years, 5, 10))
sim_n_seropositive <- get_sim_n_seropositive(
  sim_data = sim_data,
  foi = foi,
  sample_size_by_age = sample_size_by_age
)
```

`get_sim_probability` *Generate probabilities of being previously exposed to a pathogen given a historical force-of-infection.*

Description

Generate probabilities of being previously exposed to a pathogen given a historical force-of-infection.

Usage

```
get_sim_probability(sim_data, foi, mu = 0, model_type = "time-varying")
```

Arguments

`sim_data`	A dataframe object containing the following columns: `age` Age group markers `tsur` Year of the survey
`foi`	Numeric atomic vector corresponding to the desired time-varying or age-varying Force-of-Infection to simulate from
`mu`	Seroreversion rate
`model_type`	String specifying the type of model to be used. Current valid options are 'time-varying' and 'age-varying'

Value

A dataframe containing the following columns:

`age` Exposure ages

`probability` Probability to obtain a seropositive case for each age according to the provided FoI

Examples

```
n_years <- 50
sim_data <- data.frame(
  age = seq(1,n_years),
  tsur = 2050
)
foi <- rep(0.02, n_years)
sim_probability <- get_sim_probability(sim_data = sim_data, foi=foi)
```

`get_table_rhats` *Build dataframe containing the R-hat estimates for a given serological model*

Description

This function relies on rhat to extract the R-hat estimates of the serological model object `seromodel_object` and returns a table a dataframe with the estimates for each year of birth.

Usage

```
get_table_rhats(seromodel_object, cohort_ages)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`cohort_ages`	A data frame containing the age of each cohort corresponding to each birth year.

Value

rhats table

Examples

```
data(chagas2012)
serodata <- prepare_serodata(serodata = chagas2012)
model_constant <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1500
)
cohort_ages <- get_cohort_ages(serodata)
get_table_rhats(
  seromodel_object = model_constant,
  cohort_ages = cohort_ages
)
```

`group_sim_data` *Group simulated serological dataset by age*

Description

Group simulated serological dataset by age

Usage

```
group_sim_data(sim_data, col_age = "age", step = 5)
```

Arguments

`sim_data`	Dataframe with the same structure as the output of `generate_sim_data()`.
`col_age`	name of the column containing the age information
`step`	step used to split the age interval

Value

Dataframe object containing grouped simulated data generated from `foi`

`plot_foi` *Generate force-of-infection plot corresponding to the specified fitted serological model*

Description

This function generates a force-of-infection plot from the results obtained by fitting a serological model. This includes the corresponding binomial confidence interval. The x axis corresponds to the decades covered by the survey the y axis to the force-of-infection.

Usage

```
plot_foi(
  seromodel_object,
  cohort_ages,
  max_lambda = NA,
  size_text = 25,
  foi_sim = NULL
)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`cohort_ages`	A data frame containing the age of each cohort corresponding to each birth year.
`max_lambda`	TBD
`size_text`	Text size use in the theme of the graph returned by the function.
`foi_sim`	TBD

Value

A ggplot2 object containing the force-of-infection vs time including the corresponding confidence interval.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1000
)
cohort_ages <- get_cohort_ages(serodata)
plot_foi(
  seromodel_object = seromodel_object,
  cohort_ages = cohort_ages,
  size_text = 15
)
```

`plot_info_table` *Generate plot summarizing a given table*

Description

Generate plot summarizing a given table

Usage

```
plot_info_table(info_table, size_text)
```

Arguments

`info_table`	Table with the information to be summarised
`size_text`	Text size of the graph returned by the function

Value

ggplot object summarizing the information in `info_table`

Examples

```
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1000
)
seromodel_summary <- extract_seromodel_summary(
  seromodel_object = seromodel_object,
  serodata = serodata
)
info_table <- t(seromodel_summary)
plot_info_table(info_table, size_text = 15)
```

`plot_rhats` *Generate plot of the R-hat estimates for the specified fitted serological model*

Description

This function generates a plot of the R-hat estimates obtained for a specified fitted serological model `seromodel_object`. The x axis corresponds to the decades covered by the survey and the y axis to the value of the rhats. All rhats must be smaller than 1 to ensure convergence (for further details check rhat).

Usage

```
plot_rhats(seromodel_object, cohort_ages, size_text = 25)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`cohort_ages`	A data frame containing the age of each cohort corresponding to each birth year.
`size_text`	Text size use in the theme of the graph returned by the function.

Value

The rhats-convergence plot of the selected model.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1000
```

```
)
cohort_ages <- get_cohort_ages(serodata = serodata)
plot_rhats(seromodel_object,
  cohort_ages = cohort_ages,
  size_text = 15
)
```

`plot_seromodel` *Generate vertical arrangement of plots showing a summary of a model, the estimated seroprevalence, the force-of-infection fit and the R-hat estimates plots.*

Description

Generate vertical arrangement of plots showing a summary of a model, the estimated seroprevalence, the force-of-infection fit and the R-hat estimates plots.

Usage

```
plot_seromodel(
  seromodel_object,
  serodata,
  alpha = 0.05,
  max_lambda = NA,
  size_text = 25,
  bin_data = TRUE,
  bin_step = 5,
  foi_sim = NULL
)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`alpha`	Probability threshold for statistical significance used for both the binomial confidence interval, and the lower and upper quantiles of the estimated prevalence.

`max_lambda`	TBD
`size_text`	Text size use in the theme of the graph returned by the function.
`bin_data`	If `TRUE`, `serodata` is binned by means of `prepare_bin_data`. Otherwise, age groups are kept as originally input.
`bin_step`	Integer specifying the age groups bin size to be used when `bin_data` is set to `TRUE`.
`foi_sim`	TBD

Value

A ggplot object with a vertical arrange containing the seropositivity, force of infection, and convergence plots.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1000
)
plot_seromodel(seromodel_object,
  serodata = serodata,
  size_text = 15
)
```

`plot_seroprev` *Generate seropositivity plot from a raw serological survey dataset*

Description

Generate seropositivity plot from a raw serological survey dataset

Usage

```
plot_seroprev(serodata, size_text = 6, bin_data = TRUE, bin_step = 5)
```

Arguments

`serodata`	A data frame containing the data from a serological survey. This data frame must contain the following columns: `survey` survey Label of the current survey `total` Number of samples for each age group `counts` Number of positive samples for each age group `age_min` age_min

	`age_max` age_max `tsur` Year in which the survey took place `country` The country where the survey took place `test` The type of test taken `antibody` antibody Alternatively to `age_min` and `age_max`, the dataset could already include the age group marker `age_mean_f`, representing the middle point between `age_min` and `age_max`. If `afe_mean_f` is missing, it will be generated by the function.
`size_text`	Text size use in the theme of the graph returned by the function.
`bin_data`	If `TRUE`, `serodata` is binned by means of `prepare_bin_data`. Otherwise, age groups are kept as originally input.
`bin_step`	Integer specifying the age groups bin size to be used when `bin_data` is set to `TRUE`.

Value

A ggplot object containing the seropositivity-vs-age graph of the raw data of a given seroprevalence survey with its corresponding binomial confidence interval.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
plot_seroprev(serodata, size_text = 15)
```

`plot_seroprev_fitted` *Generate seropositivity plot corresponding to the specified fitted serological model*

Description

This function generates a seropositivity plot of the specified serological model object. This includes the original data grouped by age as well as the obtained fitting from the model implementation. Age is located on the x axis and seropositivity on the y axis with its corresponding confidence interval.

Usage

```
plot_seroprev_fitted(
  seromodel_object,
  serodata,
  size_text = 6,
  bin_data = TRUE,
  bin_step = 5,
  alpha = 0.05
)
```

Arguments

`seromodel_object`	Stanfit object containing the results of fitting a model by means of run_seromodel.
`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`size_text`	Text size of the graph returned by the function.
`bin_data`	If `TRUE`, `serodata` is binned by means of `prepare_bin_data`. Otherwise, age groups are kept as originally input.
`bin_step`	Integer specifying the age groups bin size to be used when `bin_data` is set to `TRUE`.
`alpha`	Probability threshold for statistical significance used for both the binomial confidence interval, and the lower and upper quantiles of the estimated prevalence.

Value

A ggplot object containing the seropositivity-vs-age graph including the data, the fitted model and their corresponding confidence intervals.

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
seromodel_object <- fit_seromodel(
  serodata = serodata,
  foi_model = "constant",
  iter = 1000
)
plot_seroprev_fitted(seromodel_object,
  serodata = serodata,
  size_text = 15
)
```

`prepare_serodata` *Prepare data from a serological survey for modelling*

Description

This function adds the necessary additional variables to the given dataset `serodata` corresponding to a serological survey.

Usage

```
prepare_serodata(serodata = serodata, alpha = 0.05)
```

Arguments

`serodata`	A data frame containing the data from a serological survey. This data frame must contain the following columns: `survey` survey Label of the current survey `total` Number of samples for each age group `counts` Number of positive samples for each age group `age_min` age_min `age_max` age_max `tsur` Year in which the survey took place `country` The country where the survey took place `test` The type of test taken `antibody` antibody Alternatively to `age_min` and `age_max`, the dataset could already include the age group marker `age_mean_f`, representing the middle point between `age_min` and `age_max`. If `afe_mean_f` is missing, it will be generated by the function.
`alpha`	probability of a type I error. For further details refer to binconf.

Value

serodata with additional columns necessary for the analysis. These columns are:

`age_mean_f`	Floor value of the average between age_min and age_max for the age groups delimited by `age_min` and `age_max`
`sample_size`	The size of the sample
`birth_year`	Years in which the subject was born according to the age group marker `age_mean_f`
`prev_obs`	Observed prevalence
`prev_obs_lower`	Lower limit of the confidence interval for the observed prevalence
`prev_obs_upper`	Upper limit of the confidence interval for the observed prevalence

Examples

```
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
```

`probability_exact_age_varying` *Returns the probability of being seropositive for age-varying force-of-infection including seroreversion*

Description

Returns the probability of being seropositive for age-varying force-of-infection including seroreversion

Usage

```
probability_exact_age_varying(age, foi, mu = 0)
```

Arguments

`age`	Integer corresponding to the age of the exposed cohort
`foi`	Numeric atomic vector corresponding to the age-varying FoI to simulate from
`mu`	Seroreversion rate

Value

probability of being seropositive for age-varying FoI including seroreversion

---

`probability_exact_time_varying` *Computes the probability of being seropositive for age-varying force-of-infection including seroreversion*

Description

Computes the probability of being seropositive for age-varying force-of-infection including seroreversion

Usage

```
probability_exact_time_varying(ages, foi, mu = 0)
```

Arguments

`ages`	Integer indicating the ages of the exposed cohorts
`foi`	Numeric atomic vector corresponding to the age-varying force-of-infection to simulate from
`mu`	Seroreversion rate

Value

probability of being seropositive for age-varying FoI including seroreversion

---

`run_seromodel` *Run specified stan model for the force-of-infection and estimate the seroprevalence based on the result of the fit*

Description

Starting on v.0.1.0, this function will be DEPRECATED. Use `fit_seromodel` instead. This function runs the specified model for the force-of-infection `foi_model` using the data from a seroprevalence survey `serodata` as the input data. See fit_seromodel for further details.

Usage

```
run_seromodel(
  serodata,
  foi_model = c("constant", "tv_normal_log", "tv_normal"),
  foi_location = 0,
  foi_scale = 1,
  chunk_size = 1,
  chunks = NULL,
  iter = 1000,
  adapt_delta = 0.9,
  max_treedepth = 10,
  seed = 12345,
  print_summary = TRUE,
  ...
)
```

Arguments

`serodata`	A data frame containing the data from a seroprevalence survey. This data frame must contain at least the following columns: `total` Number of samples for each age group `counts` Number of positive samples for each age group `tsur` Year in which the survey took place `age_mean_f` Floor value of the average between age_min and age_max `sample_size` The size of the sample `birth_year` The year in which the individuals of each age group were born The last six columns can be added to `serodata` by means of the function `prepare_serodata()`.
`foi_model`	Name of the selected model. Current version provides three options: `"constant"` Runs a constant model `"tv_normal"` Runs a normal model

	`"tv_normal_log"` Runs a normal logarithmic model
`foi_location`	Location parameter of the force-of-infection distribution of the selected model. Depending on `foi_model`, the meaning may vary.
`foi_scale`	Scale parameter of the force-of-infection distribution of the selected model. Depending on `foi_model`, the meaning may vary.
`chunk_size`	Size of the chunks to be used in case that the chunk structure `chunks` is not specified in fit_seromodel. Default is 1, meaning that one force of infection value is to be estimated for every year in the time interval spanned by the serosurvey. If the length of the time interval is not exactly divisible by `chunk_size`, the remainder years are included in the last chunk.
`chunks`	Numeric list specifying the chunk structure of the time interval from the birth year of the oldest age cohort `min(serodata$age_mean_f)` to the time when the serosurvey was conducted `t_sur`. If `NULL`, the time interval is divided in chunks of size `chunk_size`.
`iter`	Number of interactions for each chain including the warmup. `iter` in sampling.
`adapt_delta`	Real number between 0 and 1 that represents the target average acceptance probability. Increasing the value of `adapt_delta` will result in a smaller step size and fewer divergences. For further details refer to the `control` parameter in sampling or here.
`max_treedepth`	Maximum tree depth for the binary tree used in the NUTS stan sampler. For further details refer to the `control` parameter in sampling.
`seed`	For further details refer to the `seed` parameter in sampling.
`print_summary`	Boolean. If `TRUE`, a table summarizing modelling results is printed.
`...`	Additional parameters for sampling.

Value

`seromodel_object`. An object containing relevant information about the implementation of the model. For further details refer to fit_seromodel.

Examples

```
## Not run:
data(chagas2012)
serodata <- prepare_serodata(chagas2012)
run_seromodel(
  serodata,
  foi_model = "constant"
)

## End(Not run)
```

`simdata_constant` *Constant force-of-infection simulated serosurvey*

Description

Simulated dataset describing a endemic situation where the force-of-infection is constant over a period of 50 years (2000-2050) with a value of 0.2. The hypothetical serosurvey is conducted in year 2050 for 250 individuals that are up to 50 years old.

Usage

```
simdata_constant
```

Format

An object of class `"cross"`; see `qtl::read.cross()`.

Examples

```
simdata_constant
```

`simdata_large_epi` *Large epidemic simulated serosurvey*

Description

Simulated dataset describing a large epidemic between the years 2032 and 2035 with a constant force-of-infection with value 1.5. The hypothetical serosurvey is conducted in year 2050 for 250 individuals that are up to 50 years old.

Usage

```
simdata_large_epi
```

Format

An object of class `"cross"`; see `qtl::read.cross()`.

Examples

```
simdata_large_epi
```

`simdata_sw_dec` *Step-wise decreasing force-of-infection simulated serosurvey*

Description

Simulated dataset describing a situation where the force-of-infection follows a step-wise decreasing tendency over a period of 50 years (2000-2050) The hypothetical serosurvey is conducted in year 2050 for 250 individuals that are up to 50 years old.

Usage

```
simdata_sw_dec
```

Format

An object of class `"cross"`; see `qtl::read.cross()`.

Examples

```
simdata_sw_dec
```

`veev2012` *Seroprevalence data on serofoi*

Description

Data from a serological surveys

Usage

```
veev2012
```

Format

An object of class `"cross"`; see `qtl::read.cross()`.

Examples

```
veev2012
```

Index