

Package ‘IncidencePrevalence’

July 1, 2024

Title Estimate Incidence and Prevalence using the OMOP Common Data Model

Version 0.7.4

Description Calculate incidence and prevalence using data mapped to the Observational Medical Outcomes Partnership (OMOP) common data model. Incidence and prevalence can be estimated for the total population in a database or for a stratification cohort.

Encoding UTF-8

RoxygenNote 7.3.1

Depends R (>= 4.0)

Imports CDMConnector (>= 1.3.0), checkmate (>= 2.0.0), cli (>= 3.0.0), clock, dbplyr (>= 2.5.0), dplyr (>= 1.1.0), glue (>= 1.5.0), omopgenerics (>= 0.1.2), lifecycle, lubridate (>= 1.0.0), magrittr (>= 2.0.0), PatientProfiles (>= 1.1.0), purrr (>= 0.3.5), rlang (>= 1.0.0), stringr (>= 1.5.0), tidyr (>= 1.2.0), visOmopResults

Suggests knitr, rmarkdown, RPostgres, duckdb (>= 1.0.0), DBI (>= 1.0.0), odbc, here, Hmisc, epitools, tictoc, testthat (>= 0.3.1), spelling, gt, flextable, ggplot2 (>= 3.4.0), scales (>= 1.1.0)

Config/testthat/edition 3

Config/testthat/parallel true

VignetteBuilder knitr

Language en-US

License Apache License (>= 2)

URL <https://darwin-eu.github.io/IncidencePrevalence/>

NeedsCompilation no

Author Edward Burn [aut, cre] (<<https://orcid.org/0000-0002-9286-1128>>),
Berta Raventos [aut] (<<https://orcid.org/0000-0002-4668-2970>>),
Marti Catala [aut] (<<https://orcid.org/0000-0003-3308-9905>>),
Mike Du [ctb] (<<https://orcid.org/0000-0002-9517-8834>>),
Yuchen Guo [ctb] (<<https://orcid.org/0000-0002-0847-4855>>),

Adam Black [ctb] (<<https://orcid.org/0000-0001-5576-8701>>),
 Ger Inberg [ctb] (<<https://orcid.org/0000-0001-8993-8748>>),
 Kim Lopez [ctb] (<<https://orcid.org/0000-0002-8462-8668>>)

Maintainer Edward Burn <edward.burn@dorms.ox.ac.uk>

Repository CRAN

Date/Publication 2024-07-01 20:50:02 UTC

Contents

benchmarkIncidencePrevalence	2
estimateIncidence	3
estimatePeriodPrevalence	5
estimatePointPrevalence	7
generateDenominatorCohortSet	8
generateTargetDenominatorCohortSet	10
mockIncidencePrevalenceRef	11
optionsTableIncidence	13
optionsTablePrevalence	13
participants	14
plotIncidence	15
plotPrevalence	16
Index	18

benchmarkIncidencePrevalence

Run benchmark of incidence and prevalence analyses

Description

Run benchmark of incidence and prevalence analyses

Usage

```
benchmarkIncidencePrevalence(  
  cdm,  
  returnParticipants = FALSE,  
  analysisType = "all"  
)
```

Arguments

cdm	A CDM reference object
returnParticipants	Whether to return participants
analysisType	A string of the following: "all", "only incidence", "only prevalence"

Value

a tibble with time taken for different analyses

Examples

```
cdm <- mockIncidencePrevalenceRef(
  sampleSize = 100,
  earliestObservationStartDate = as.Date("2010-01-01"),
  latestObservationStartDate = as.Date("2010-01-01"),
  minDaysToObservationEnd = 364,
  maxDaysToObservationEnd = 364,
  outPre = 0.1
)

timings <- benchmarkIncidencePrevalence(cdm)
```

estimateIncidence	<i>Collect population incidence estimates</i>
-------------------	---

Description

Collect population incidence estimates

Usage

```
estimateIncidence(
  cdm,
  denominatorTable,
  outcomeTable,
  denominatorCohortId = NULL,
  outcomeCohortId = NULL,
  interval = "years",
  completeDatabaseIntervals = TRUE,
  outcomeWashout = Inf,
  repeatedEvents = FALSE,
  minCellCount = 5,
  strata = list(),
  includeOverallStrata = TRUE,
  returnParticipants = FALSE
)
```

Arguments

cdm	A CDM reference object
denominatorTable	A cohort table with a set of denominator cohorts (for example, created using the generateDenominatorCohortSet() function).

outcomeTable	A cohort table in the cdm reference containing a set of outcome cohorts.
denominatorCohortId	The cohort definition ids of the denominator cohorts of interest. If NULL all cohorts will be considered in the analysis.
outcomeCohortId	The cohort definition ids of the outcome cohorts of interest. If NULL all cohorts will be considered in the analysis.
interval	Time intervals over which incidence is estimated. Can be "weeks", "months", "quarters", "years", or "overall". ISO weeks will be used for weeks. Calendar months, quarters, or years can be used, or an overall estimate for the entire time period observed (from earliest cohort start to last cohort end) can also be estimated. If more than one option is chosen then results will be estimated for each chosen interval.
completeDatabaseIntervals	TRUE/ FALSE. Where TRUE, incidence will only be estimated for those intervals where the denominator cohort captures all the interval.
outcomeWashout	The number of days used for a 'washout' period between the end of one outcome and an individual starting to contribute time at risk. If Inf, no time can be contributed after an event has occurred.
repeatedEvents	TRUE/ FALSE. If TRUE, an individual will be able to contribute multiple events during the study period (time while they are present in an outcome cohort and any subsequent washout will be excluded). If FALSE, an individual will only contribute time up to their first event.
minCellCount	The minimum number of events to reported, below which results will be obscured. If 0, all results will be reported.
strata	Variables added to the denominator cohort table for which to stratify estimates.
includeOverallStrata	Whether to include an overall result as well as strata specific results (when strata has been specified).
returnParticipants	Either TRUE or FALSE. If TRUE references to participants from the analysis will be returned allowing for further analysis. Note, if using permanent tables and returnParticipants is TRUE, one table per analysis will be kept in the cdm write schema.

Value

Incidence estimates

Examples

```

cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator",
  cohortDateRange = c(as.Date("2008-01-01"), as.Date("2018-01-01"))
)
inc <- estimateIncidence(

```

```
    cdm = cdm,  
    denominatorTable = "denominator",  
    outcomeTable = "outcome"  
  )
```

estimatePeriodPrevalence

Estimate period prevalence

Description

Estimate period prevalence

Usage

```
estimatePeriodPrevalence(  
  cdm,  
  denominatorTable,  
  outcomeTable,  
  denominatorCohortId = NULL,  
  outcomeCohortId = NULL,  
  interval = "years",  
  completeDatabaseIntervals = TRUE,  
  fullContribution = FALSE,  
  strata = list(),  
  includeOverallStrata = TRUE,  
  minCellCount = 5,  
  returnParticipants = FALSE  
)
```

Arguments

cdm	A CDM reference object
denominatorTable	A cohort table with a set of denominator cohorts (for example, created using the generateDenominatorCohortSet() function).
outcomeTable	A cohort table in the cdm reference containing a set of outcome cohorts.
denominatorCohortId	The cohort definition ids of the denominator cohorts of interest. If NULL all cohorts will be considered in the analysis.
outcomeCohortId	The cohort definition ids of the outcome cohorts of interest. If NULL all cohorts will be considered in the analysis.

<code>interval</code>	Time intervals over which period prevalence is estimated. This can be "weeks", "months", "quarters", "years", or "overall". ISO weeks will be used for weeks. Calendar months, quarters, or years can be used as the period. If more than one option is chosen then results will be estimated for each chosen interval.
<code>completeDatabaseIntervals</code>	TRUE/ FALSE. Where TRUE, prevalence will only be estimated for those intervals where the database captures all the interval (based on the earliest and latest observation period start dates, respectively).
<code>fullContribution</code>	TRUE/ FALSE. Where TRUE, individuals will only be included if they in the database for the entire interval of interest. If FALSE they are only required to present for one day of the interval in order to contribute.
<code>strata</code>	Variables added to the denominator cohort table for which to stratify estimates.
<code>includeOverallStrata</code>	Whether to include an overall result as well as strata specific results (when strata has been specified).
<code>minCellCount</code>	Minimum number of events to report- results lower than this will be obscured. If NULL all results will be reported.
<code>returnParticipants</code>	Either TRUE or FALSE. If TRUE references to participants from the analysis will be returned allowing for further analysis. Note, if using permanent tables and <code>returnParticipants</code> is TRUE, one table per analysis will be kept in the <code>cdm</code> write schema.

Value

Period prevalence estimates

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator",
  cohortDateRange = c(as.Date("2008-01-01"), as.Date("2018-01-01"))
)
estimatePeriodPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome",
  interval = "months"
)
```

estimatePointPrevalence
Estimate point prevalence

Description

Estimate point prevalence

Usage

```
estimatePointPrevalence(  
  cdm,  
  denominatorTable,  
  outcomeTable,  
  denominatorCohortId = NULL,  
  outcomeCohortId = NULL,  
  interval = "years",  
  timePoint = "start",  
  strata = list(),  
  includeOverallStrata = TRUE,  
  minCellCount = 5,  
  returnParticipants = FALSE  
)
```

Arguments

cdm	A CDM reference object
denominatorTable	A cohort table with a set of denominator cohorts (for example, created using the generateDenominatorCohortSet() function).
outcomeTable	A cohort table in the cdm reference containing a set of outcome cohorts.
denominatorCohortId	The cohort definition ids of the denominator cohorts of interest. If NULL all cohorts will be considered in the analysis.
outcomeCohortId	The cohort definition ids of the outcome cohorts of interest. If NULL all cohorts will be considered in the analysis.
interval	Time intervals over which period prevalence is estimated. Can be "weeks", "months", "quarters", or "years". ISO weeks will be used for weeks. Calendar months, quarters, or years can be used as the period. If more than one option is chosen then results will be estimated for each chosen interval.
timePoint	where to compute the point prevalence
strata	Variables added to the denominator cohort table for which to stratify estimates.
includeOverallStrata	Whether to include an overall result as well as strata specific results (when strata has been specified).

`minCellCount` Minimum number of events to report- results lower than this will be obscured. If NULL all results will be reported.

`returnParticipants` Either TRUE or FALSE. If TRUE references to participants from the analysis will be returned allowing for further analysis. Note, if using permanent tables and `returnParticipants` is TRUE, one table per analysis will be kept in the `cdm` write schema.

Value

Point prevalence estimates

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator",
  cohortDateRange = c(as.Date("2008-01-01"), as.Date("2018-01-01"))
)
estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome",
  interval = "months"
)
```

`generateDenominatorCohortSet`

Identify a set of denominator populations

Description

`generateDenominatorCohortSet()` creates a set of cohorts that can be used for the denominator population in analyses of incidence, using `estimateIncidence()`, or prevalence, using `estimatePointPrevalence()` or `estimatePeriodPrevalence()`.

Usage

```
generateDenominatorCohortSet(
  cdm,
  name,
  cohortDateRange = as.Date(c(NA, NA)),
  ageGroup = list(c(0, 150)),
  sex = "Both",
  daysPriorObservation = 0,
  requirementInteractions = TRUE
)
```


Arguments

cdm	A CDM reference object
name	Name of the cohort table to be created. Note if a table already exists with this name in the database (give the prefix being used for the cdm reference) it will be overwritten.
cohortDateRange	Two dates. The first indicating the earliest cohort start date and the second indicating the latest possible cohort end date. If NULL or the first date is set as missing, the earliest observation_start_date in the observation_period table will be used for the former. If NULL or the second date is set as missing, the latest observation_end_date in the observation_period table will be used for the latter.
ageGroup	A list of age groups for which cohorts will be generated. A value of <code>list(c(0, 17), c(18, 30))</code> would, for example, lead to the creation of cohorts for those aged from 0 to 17, and from 18 to 30. In this example an individual turning 18 during the time period would appear in both cohorts (leaving the first cohort the day before their 18th birthday and entering the second from the day of their 18th birthday).
sex	Sex of the cohorts. This can be one or more of: "Male", "Female", or "Both".
daysPriorObservation	The number of days of prior observation observed in the database required for an individual to start contributing time in a cohort.
requirementInteractions	If TRUE, cohorts will be created for all combinations of ageGroup, sex, and daysPriorObservation. If FALSE, only the first value specified for the other factors will be used. Consequently, order of values matters when requirementInteractions is FALSE.

Value

A cdm reference

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm,
  name = "denominator",
  cohortDateRange = as.Date(c("2008-01-01", "2020-01-01"))
)
cdm
```

```
generateTargetDenominatorCohortSet
```

Identify a set of denominator populations using a target cohort

Description

generateTargetDenominatorCohortSet() creates a set of cohorts that can be used for the denominator population in analyses of incidence, using estimateIncidence(), or prevalence, using estimatePointPrevalence() or estimatePeriodPrevalence().

Usage

```
generateTargetDenominatorCohortSet(
  cdm,
  name,
  targetCohortTable,
  targetCohortId = NULL,
  cohortDateRange = as.Date(c(NA, NA)),
  ageGroup = list(c(0, 150)),
  sex = "Both",
  daysPriorObservation = 0,
  requirementInteractions = TRUE
)
```

Arguments

cdm	A CDM reference object
name	Name of the cohort table to be created.
targetCohortTable	A cohort table in the cdm reference to use to limit cohort entry and exit (with individuals only contributing to a cohort when they are contributing to the cohort in the target table).
targetCohortId	The cohort definition id for the cohort of interest in the target table. If targetCohortTable is specified, a single targetCohortId must also be specified.
cohortDateRange	Two dates. The first indicating the earliest cohort start date and the second indicating the latest possible cohort end date. If NULL or the first date is set as missing, the earliest observation_start_date in the observation_period table will be used for the former. If NULL or the second date is set as missing, the latest observation_end_date in the observation_period table will be used for the latter.
ageGroup	A list of age groups for which cohorts will be generated. A value of list(c(0, 17), c(18, 30)) would, for example, lead to the creation of cohorts for those aged from 0 to 17, and from 18 to 30. In this example an individual turning 18 during the time period would appear in both cohorts (leaving the first cohort the day before their 18th birthday and entering the second from the day of their 18th birthday).

sex Sex of the cohorts. This can be one or more of: "Male", "Female", or "Both".

daysPriorObservation The number of days of prior observation observed in the database required for an individual to start contributing time in a cohort.

requirementInteractions If TRUE, cohorts will be created for all combinations of ageGroup, sex, and daysPriorObservation. If FALSE, only the first value specified for the other factors will be used. Consequently, order of values matters when requirementInteractions is FALSE.

Value

A cdm reference

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateTargetDenominatorCohortSet(
  cdm = cdm,
  name = "denominator",
  targetCohortTable = "target",
  cohortDateRange = as.Date(c("2008-01-01", "2020-01-01"))
)
cdm
```

```
mockIncidencePrevalenceRef
```

Generate example subset of the OMOP CDM for estimating incidence and prevalence

Description

Generate example subset of the OMOP CDM for estimating incidence and prevalence

Usage

```
mockIncidencePrevalenceRef(
  personTable = NULL,
  observationPeriodTable = NULL,
  targetCohortTable = NULL,
  outcomeTable = NULL,
  sampleSize = 1,
  outPre = 1,
  seed = 444,
  ageBeta = NULL,
  genderBeta = NULL,
  intercept = NULL,
```

```

earliestDateOfBirth = NULL,
latestDateOfBirth = NULL,
earliestObservationStartDate = NULL,
latestObservationStartDate = NULL,
minDaysToObservationEnd = NULL,
maxDaysToObservationEnd = NULL,
minOutcomeDays = 1,
maxOutcomeDays = 10,
maxOutcomes = 1
)

```

Arguments

personTable A tibble in the format of the person table.

observationPeriodTable A tibble in the format of the observation period table.

targetCohortTable A tibble in the format of a cohort table which can be used for stratification

outcomeTable A tibble in the format of a cohort table which can be used for outcomes

sampleSize The number of unique patients.

outPre The fraction of patients with an event.

seed The seed for simulating the data set. Use the same seed to get same data set.

ageBeta The beta for the standardised age in a logistic regression outcome model.

genderBeta The beta for the gender flag in a logistic regression outcome model.

intercept The beta for the intercept in a logistic regression outcome model.

earliestDateOfBirth The earliest date of birth of a patient in person table.

latestDateOfBirth The latest date of birth of a patient in person table.

earliestObservationStartDate The earliest observation start date for patient format.

latestObservationStartDate The latest observation start date for patient format.

minDaysToObservationEnd The minimum number of days of the observational integer.

maxDaysToObservationEnd The maximum number of days of the observation period integer.

minOutcomeDays The minimum number of days of the outcome period default set to 1.

maxOutcomeDays The maximum number of days of the outcome period default set to 10.

maxOutcomes The maximum possible number of outcomes per person can have default set to 1.

Value

A cdm reference to a duckdb database with mock data.

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 100)
cdm
```

optionsTableIncidence *Additional arguments for the functions tableIncidence.*

Description

It provides a list of allowed inputs for .option argument in tableIncidence, and their given default values.

Usage

```
optionsTableIncidence()
```

Value

The default .options named list.

Examples

```
{
  optionsTableIncidence()
}
```

optionsTablePrevalence
Additional arguments for the functions tablePrevalence.

Description

It provides a list of allowed inputs for .option argument in tablePrevalence, and their given default values.

Usage

```
optionsTablePrevalence()
```

Value

The default .options named list.

Examples

```
{
  optionsTablePrevalence()
}
```

participants

Participants contributing to an analysis

Description

Participants contributing to an analysis

Usage

```
participants(result, analysisId)
```

Arguments

result	Result object
analysisId	ID of a specific analysis to return participants for

Value

References to tables with the study participants contributing to a given analysis

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 200)
cdm <- generateDenominatorCohortSet(cdm, name = "denominator")
incidence <- estimateIncidence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome",
  interval = "overall"
)
participants(result = incidence, analysisId = 1)
```

plotIncidence	<i>Plot incidence results</i>
---------------	-------------------------------

Description

Plot incidence results

Usage

```
plotIncidence(  
  result,  
  x = "incidence_start_date",  
  ylim = c(0, NA),  
  ribbon = FALSE,  
  facet = NULL,  
  colour = NULL,  
  colour_name = NULL,  
  options = list()  
)
```

Arguments

result	Incidence results
x	Variable to plot on x axis
ylim	Limits for the Y axis
ribbon	If TRUE, the plot will join points using a ribbon
facet	Variables to use for facets
colour	Variables to use for colours
colour_name	Colour legend name
options	a list of optional plot options

Value

A ggplot with the incidence results plotted

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)  
cdm <- generateDenominatorCohortSet(  
  cdm = cdm, name = "denominator",  
  cohortDateRange = c(as.Date("2008-01-01"), as.Date("2018-01-01"))  
)  
inc <- estimateIncidence(  
  cdm = cdm,  
  denominatorTable = "denominator",  
  outcomeTable = "outcome")
```

```
)  
plotIncidence(inc)
```

plotPrevalence	<i>Plot prevalence results</i>
----------------	--------------------------------

Description

Plot prevalence results

Usage

```
plotPrevalence(  
  result,  
  x = "prevalence_start_date",  
  ylim = c(0, NA),  
  ribbon = FALSE,  
  facet = NULL,  
  colour = NULL,  
  colour_name = NULL,  
  options = list()  
)
```

Arguments

result	Prevalence results
x	Variable to plot on x axis
ylim	Limits for the Y axis
ribbon	If TRUE, the plot will join points using a ribbon
facet	Variables to use for facets
colour	Variables to use for colours
colour_name	Colour legend name
options	a list of optional plot options

Value

A ggplot with the prevalence results plotted

Examples

```
cdm <- mockIncidencePrevalenceRef(sampleSize = 1000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator",
  cohortDateRange = c(as.Date("2014-01-01"), as.Date("2018-01-01"))
)
prev <- estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
plotPrevalence(prev)
```

Index

`benchmarkIncidencePrevalence`, [2](#)

`estimateIncidence`, [3](#)

`estimatePeriodPrevalence`, [5](#)

`estimatePointPrevalence`, [7](#)

`generateDenominatorCohortSet`, [8](#)

`generateTargetDenominatorCohortSet`, [10](#)

`mockIncidencePrevalenceRef`, [11](#)

`optionsTableIncidence`, [13](#)

`optionsTablePrevalence`, [13](#)

`participants`, [14](#)

`plotIncidence`, [15](#)

`plotPrevalence`, [16](#)