Package ‘IncidencePrevalence’

July 11, 2023

Title Estimate Incidence and Prevalence using the OMOP Common Data Model

Version 0.4.1

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 7.2.3

Depends R (>= 4.0)

Imports CDMConnector (>= 1.0.0), checkmate (>= 2.0.0), cli (>= 3.0.0), DBI (>= 1.0.0), dbplyr (>= 2.0.0), dplyr (>= 1.1.0), glue (>= 1.5.0), ggplot2 (>= 3.4.0), scales (>= 1.1.0), lubridate (>= 1.0.0), magrittr (>= 2.0.0), purrr (>= 0.3.5), rlang (>= 1.0.0), stringr (>= 1.5.0), tidyselect (>= 1.2.0), zip (>= 2.2.0)

Suggests knitr, rmarkdown, RPostgres, tibble, duckdb, odbc, here, Hmisc, epitools, tictoc, testthat (>= 0.3.1), spelling, PaRe

Config/testthat/edition 3

VignetteBuilder knitr

Language en-US

License Apache License (>= 2)


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)
benchmarkIncidencePrevalence

Run benchmark of incidence and prevalence analyses

Description

Run benchmark of incidence and prevalence analyses

Usage

```
benchmarkIncidencePrevalence(
    cdm,
    cohortDateRange = NULL,
    temporary = TRUE,
    returnParticipants = FALSE,
    nOutcomes = 1,
    prevOutcomes = 0.25,
    analysisType = "all",
    outputFolder = NULL,
    fileName = NULL
)
```
**bindIncidenceEstimates**

**Arguments**

- **cdm**: A CDM reference object
- **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.
- **temporary**: If TRUE, temporary tables will be used throughout. If FALSE, permanent tables will be created in the write_schema of the cdm using the write_prefix (if specified). Note existing permanent tables in the write schema starting with the write_prefix will be at risk of being dropped or overwritten.
- **returnParticipants**: Whether to return participants (requires temporary to be FALSE)
- **nOutcomes**: An integer specifying the number of outcomes to create in the denominator cohort
- **prevOutcomes**: An array of integers for the prevalence of the outcomes in the population (in %). If the user wants all the outcomes with the same prevalence, they can also provide a single integer
- **analysisType**: A string of the following: "all", "only incidence", "only prevalence"
- **outputFolder**: Folder to save results as CSV
- **fileName**: Name given to CSV with results

**Value**

A tibble with time taken for different analyses

**Examples**

```r
cdm <- mockIncidencePrevalenceRef(
  sampleSize = 10000,
  earliestObservationStartDate = as.Date("2010-01-01"),
  latestObservationStartDate = as.Date("2018-01-01")
)
timings <- IncidencePrevalence::benchmarkIncidencePrevalence(cdm)
```

---

**bindIncidenceEstimates**

*Bind multiple incidence estimates into a single set of estimates*

**Description**

Bind multiple incidence estimates into a single set of estimates
bindPrevalenceEstimates

Usage

bindPrevalenceEstimates(...)

Arguments

... Multiple prevalence estimates, generated from estimatePeriodPrevalence() or estimatePointPrevalence()

Value

Bound prevalence estimates

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator"
)
inc1 <- estimateIncidence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
inc2 <- estimateIncidence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
incCombined <- bindIncidenceEstimates(inc1, inc2)
estimateIncidence

Value

Bound prevalence estimates

Examples

```r
cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator"
)
prev1 <- estimatePeriodPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
prev2 <- estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
prevCombined <- bindPrevalenceEstimates(prev1, prev2)
```

---

**estimateIncidence**  
*Collect population incidence estimates*

**Description**  
Collect population incidence estimates

**Usage**

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

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 inter-
vals where the database captures all the interval (based on the earliest and latest
observation period start dates, respectively).

outcomeWashout
The number of days used for a 'washout' period between the end of one out-
come and an individual starting to contribute time at risk. If Inf, no time can be
contributed after an event has occurred (whether during the study period or if
occurring beforehand).

.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 during the study period.

.minCellCount
The minimum number of events to reported, below which results will be ob-
scured. If 0, all results will be reported.

temporary
If TRUE, temporary tables will be used throughout. If FALSE, permanent ta-
bles will be created in the write_schema of the cdm using the write_prefix (if
specified). Note existing permanent tables in the write schema starting with the
write_prefix will be at risk of being dropped or overwritten.

.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
**estimatePeriodPrevalence**

*Estimate period prevalence*

**Description**

Estimate period prevalence

**Usage**

```r
estimatePeriodPrevalence(
  cdm,
  denominatorTable,
  outcomeTable,
  denominatorCohortId = NULL,
  outcomeCohortId = NULL,
  outcomeLookbackDays = 0,
  interval = "years",
  completeDatabaseIntervals = TRUE,
  fullContribution = FALSE,
  minCellCount = 5,
  temporary = 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.

**Examples**

```r
cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
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

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.

outcomeLookbackDays
Days lookback when considering an outcome as prevalent. If NULL any prior outcome will be considered as prevalent. If 0, only ongoing outcomes will be considered as prevalent.

interval
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.

completeDatabaseIntervals
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).

fullContribution
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.

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

temporary
If TRUE, temporary tables will be used throughout. If FALSE, permanent tables will be created in the write_schema of the cdm using the write_prefix (if specified). Note existing permanent tables in the write schema starting with the write_prefix will be at risk of being dropped or overwritten.

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
Period prevalence estimates

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
cdm <- generateDenominatorCohortSet(
    cdm = cdm, name = "denominator",
    cohortDateRange = c(as.Date("2008-01-01"), as.Date("2018-01-01"))
)
estimatePeriodPrevalence(
    cdm = cdm,
estimatePointPrevalence

Estimate point prevalence

Description

Estimate point prevalence

Usage

estimatePointPrevalence(
  cdm,
  denominatorTable,
  outcomeTable,
  denominatorCohortId = NULL,
  outcomeCohortId = NULL,
  outcomeLookbackDays = 0,
  interval = "years",
  timePoint = "start",
  minCellCount = 5,
  temporary = 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.
outcomeLookbackDays Days lookback when considering an outcome as prevalent. If NULL any prior outcome will be considered as prevalent. If 0, only ongoing outcomes will be considered as prevalent.
exportIncidencePrevalenceResults

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

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

temporary
If TRUE, temporary tables will be used throughout. If FALSE, permanent tables will be created in the write_schema of the cdm using the write_prefix (if specified). Note existing permanent tables in the write schema starting with the write_prefix will be at risk of being dropped or overwritten.

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

```r
cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
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"
)
```

exportIncidencePrevalenceResults

Export IncidencePrevalence results

Description
Export IncidencePrevalence results

Usage

`exportIncidencePrevalenceResults(resultList, zipName, outputFolder)`
generateDenominatorCohortSet

Arguments

- resultList: Named list with results from estimateIncidence, estimatePointPrevalence, or estimatePeriodPrevalence
- zipName: name to give zip folder
- outputFolder: directory to save zip folder containing results as a set of CSV files

Value

zip folder of results saved in the outputFolder

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
cdm <- generateDenominatorCohortSet(
  cdm = cdm, name = "denominator"
)
prev <- estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
exportIncidencePrevalenceResults(
  resultList = list("prevalence" = prev),
  zipName = "test",
  outputFolder = tempdir()
)

---

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 = "denominator",
  cohortDateRange = NULL,
  ageGroup = list(c(0, 150)),
  sex = "Both",
)
generateDenominatorCohortSet

daysPriorHistory = 0,
requirementInteractions = TRUE,
strataTable = NULL,
strataCohortId = NULL,
closedCohort = FALSE,
temporary = TRUE
)

Arguments

cdm 
A CDM reference object

name 
Name of the cohort table to be created.

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 (up to the day before their 18th birthday), and from 18 (starting the day of their 18th birthday) to 30 (up to the day before their 31st birthday).

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

daysPriorHistory 
The number of days of prior history 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 daysPriorHistory. If FALSE, only the first value specified for the other factors will be used. Consequently, order of values matters when requirementInteractions is FALSE.

strataTable 
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 strata table).

strataCohortId 
The cohort definition id for the cohort of interest in the strata table. If strataTable is specified, a single strataCohortId must also be specified.

closedCohort 
If TRUE, a closed cohort will be defined where only those individuals satisfying eligibility criteria on the start date given in cohortDateRange are included.

temporary 
If TRUE, temporary tables will be used throughout. If FALSE, permanent tables will be created in the write_schema of the cdm using the write_prefix (if specified). Note existing permanent tables in the write schema starting with the write_prefix will be at risk of being dropped or overwritten.

Value

A cohort reference
Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
cdm$denominator <- generateDenominatorCohortSet(
  cdm = cdm,
  cohortDateRange = as.Date(c("2008-01-01", "2018-01-01"))
)

description

## incidenceAttrition

Attrition associated with an incidence analysis

### Usage

incidenceAttrition(result)

### Arguments

- **result**
  
  Result for which to get attrition

### Value

tibble with counts and reasons for attrition.

Examples

```r

cdm <- mockIncidencePrevalenceRef(sampleSize = 200)
cdm <- generateDenominatorCohortSet(cdm, name = "denominator")
inc <- estimateIncidence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome",
  interval = "overall"
)
incidenceAttrition(inc)
```
description

Settings associated with an incidence analysis

Usage

incidenceSet(result)

Arguments

result  Result for which to get settings

Value

tibble with settings used when estimating incidence

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 200)
cdm <- generateDenominatorCohortSet(cdm, name = "denominator")
ing <- estimateIncidence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome",
  interval = "overall"
)
incidenceSet(inc)

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
mockIncidencePrevalenceRef

Usage

mockIncidencePrevalenceRef(
  personTable = NULL,
  observationPeriodTable = NULL,
  strataTable = 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.
strataTable  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

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

---

<table>
<thead>
<tr>
<th>participants</th>
<th>Participants contributing to an analysis</th>
</tr>
</thead>
</table>

Description

Participants contributing to an analysis

Usage

```r
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

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

Description

Plot incidence results

Usage

```r
plotIncidence(
  result,
  x = "incidence_start_date",
  ylim = c(0, NA),
  ribbon = TRUE,
  facet = NULL,
  colour = NULL,
  colour_name = NULL
)
```

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

Value

A ggplot with the incidence results plotted
Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
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  Plot prevalence results

Description

Plot prevalence results

Usage

plotPrevalence(
  result,
  x = "prevalence_start_date",
  ylim = c(0, NA),
  ribbon = TRUE,
  facet = NULL,
  colour = NULL,
  colour_name = NULL
)

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

Value

A ggplot with the prevalence results plotted
Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 10000)
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)

prevalenceAttrition  Attrition associated with an prevalence analysis

Description

Attrition associated with an prevalence analysis

Usage

prevalenceAttrition(result)

Arguments

result  Result for which to get attrition

Value

tibble with counts and reasons for attrition.

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 200)
cdm <- generateDenominatorCohortSet(cdm, name = "denominator")
prev <- estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
prevalenceAttrition(prev)
prevalenceSet

Prevalence analysis settings

Description

Settings associated with a prevalence analysis

Usage

prevalenceSet(result)

Arguments

result  
Result for which to get settings

Value

tibble with settings used when estimating prevalence

Examples

cdm <- mockIncidencePrevalenceRef(sampleSize = 200)
cdm <- generateDenominatorCohortSet(cdm, name = "denominator")prev <- estimatePointPrevalence(
  cdm = cdm,
  denominatorTable = "denominator",
  outcomeTable = "outcome"
)
prevalenceSet(prev)
Index

benchmarkIncidencePrevalence, 2
bindIncidenceEstimates, 3
bindPrevalenceEstimates, 4

estimateIncidence, 5
estimatePeriodPrevalence, 7
estimatePointPrevalence, 9
exportIncidencePrevalenceResults, 10

generateDenominatorCohortSet, 11

incidenceAttrition, 13
incidenceSet, 14

mockIncidencePrevalenceRef, 14

participants, 16
plotIncidence, 17
plotPrevalence, 18
prevalenceAttrition, 19
prevalenceSet, 20