IncidencePrevalence R Package: Use and development

Berta Raventós

Abril 2024

INCIDENCE
PREVALENCE

Edward Burn

Martí Català

Berta Raventós

Mike Du

Yuchen Guo

Adam Black

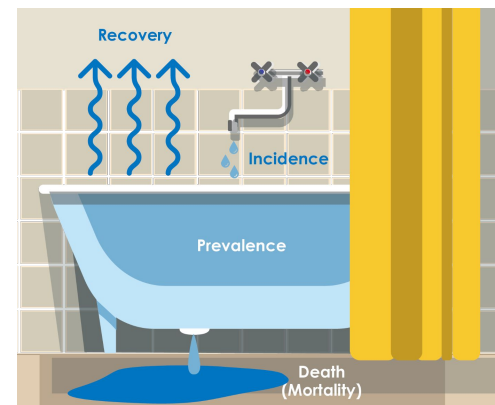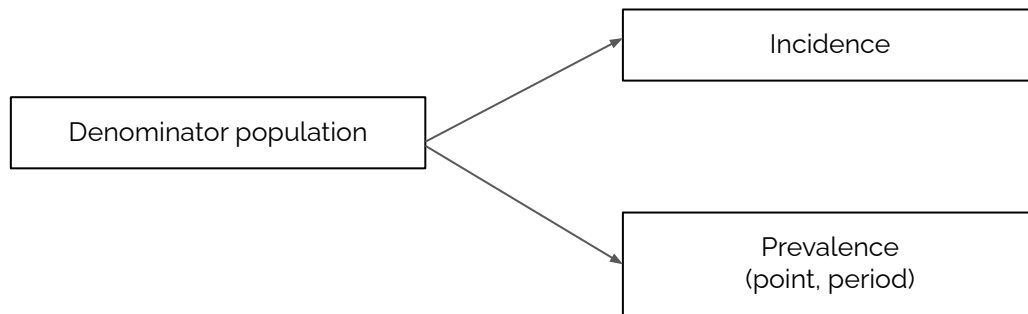Ger Inberg

Kim López

Contents

1. **Concepts**

2. Interface

3. Package Development
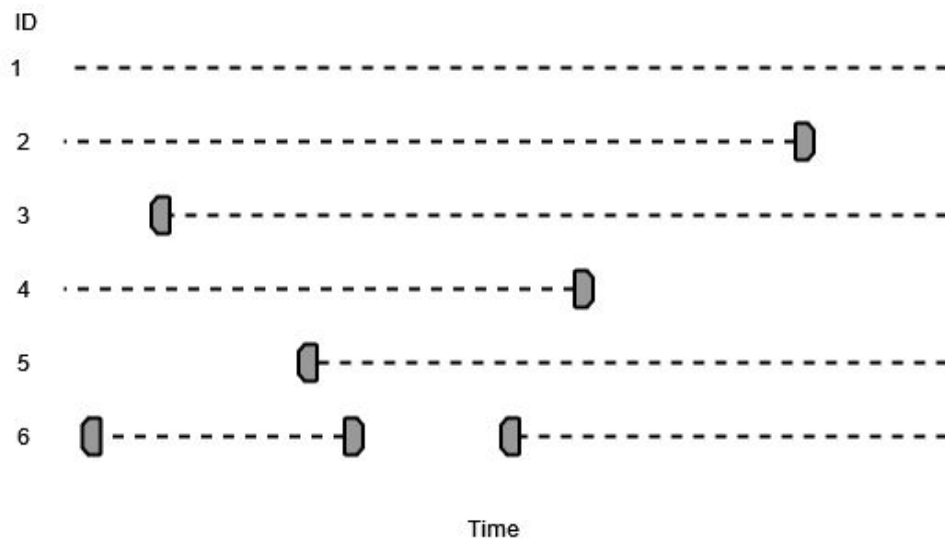
   a. Validation studies

   b. Benchmarking

- Support the analyses of population-level incidence and prevalence in datasets mapped to the OMOP CDM.

- Flexible enough to accommodate different requirements.

Observation periods

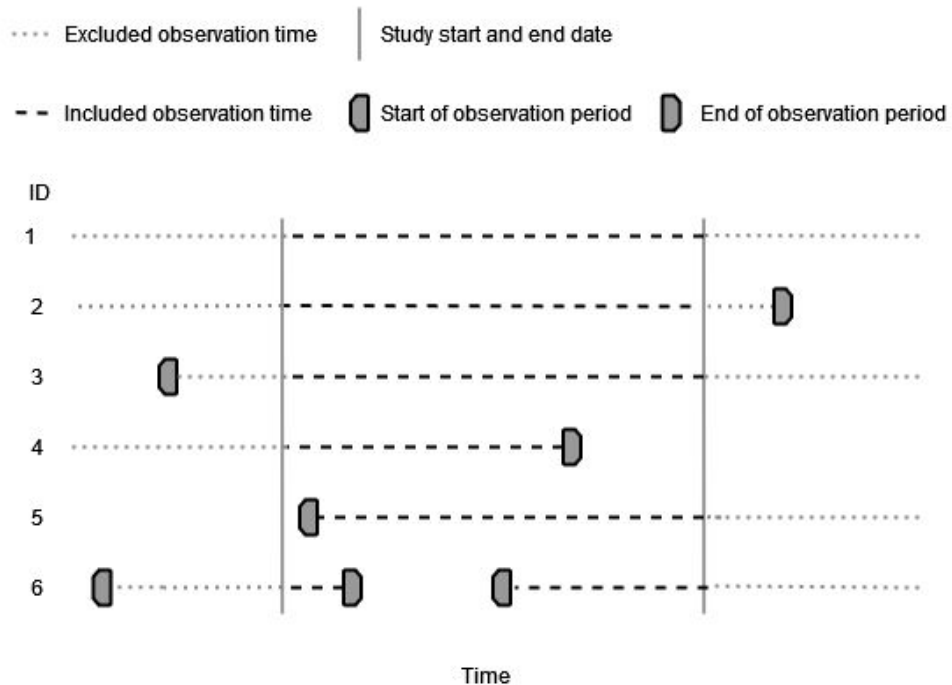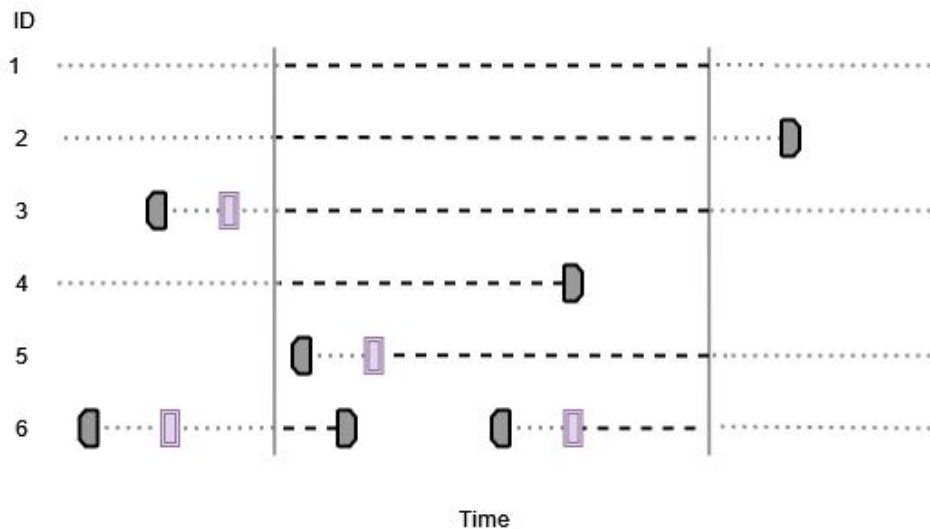Observation periods + study periods

Denominator population

Observation periods + study periods + prior history requirement

7

Observation periods + study periods + prior history requirement + age and sex restrictions

Denominator population

Observation periods + study periods + prior history requirement + age and sex restrictions

···· Excluded observation time	Study start and end date	Sufficient prior history
– – Included observation time	Start of observation period	End of observation period
Minimum age	Maximum age	
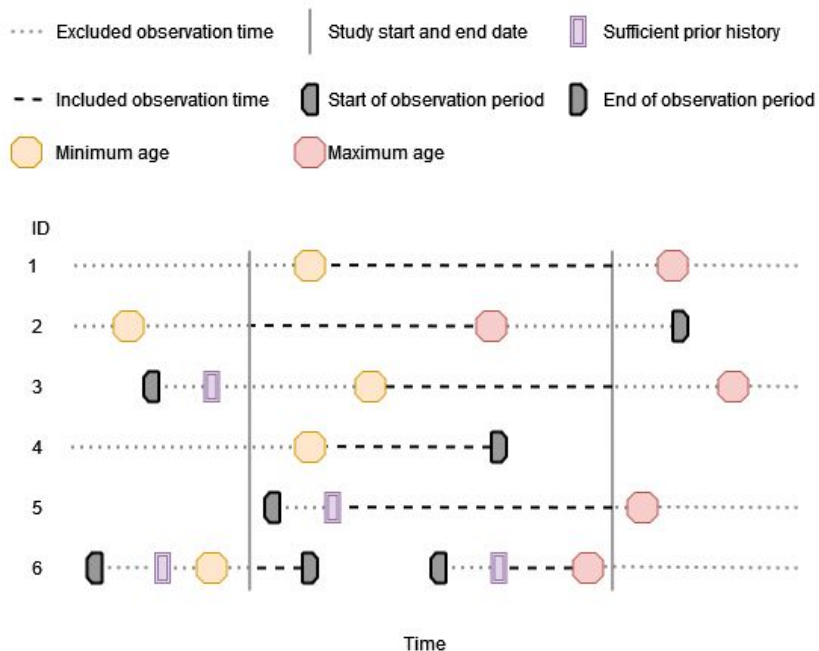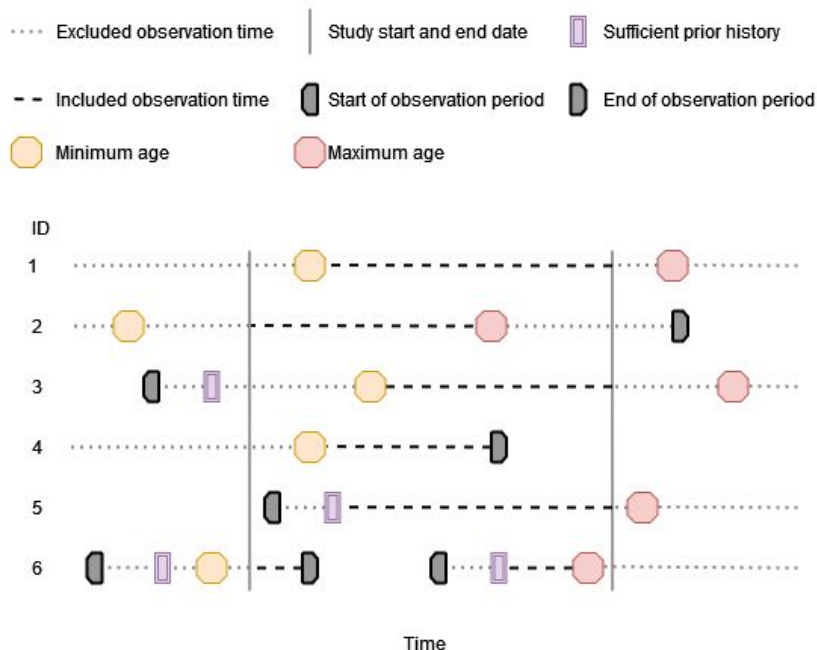
Enter the denominator population (latest of):

- Study start date
- Date at which they have sufficient prior history
- Date at which they reach a minimum age

Exit the denominator population (earliest of):

- Study end date
- Date at which their observation period ends
- The last day in which they have the maximum age

Observation periods + study periods + prior history requirement + age and sex restrictions + target cohorts
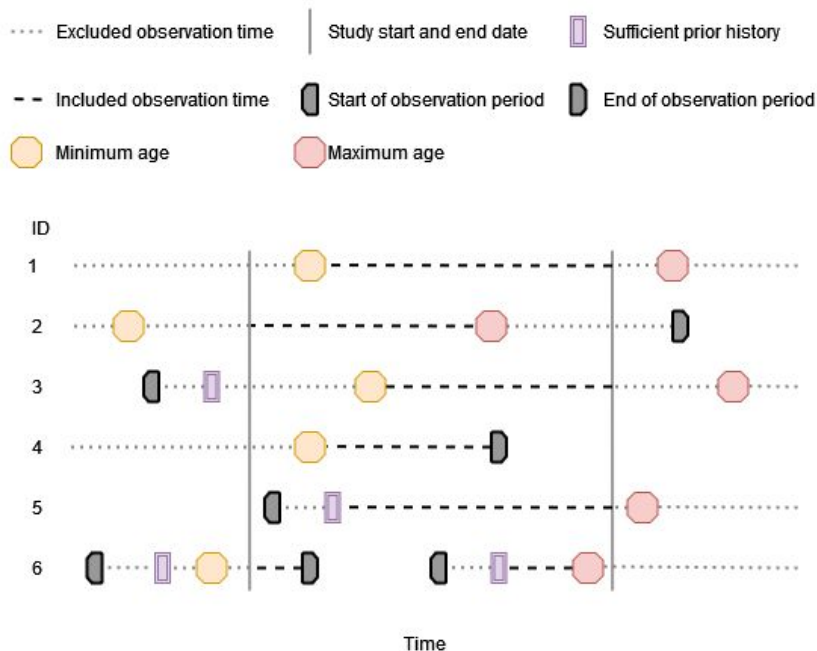
Enter the denominator population (latest of):

- Study start date
- Date at which they have sufficient prior history
- Date at which they reach a minimum age
- Date at which they enter the target cohort

Exit the denominator population (earliest of):

- Study end date
- Date at which their observation period ends
- The last day in which they have the maximum age
- Date at which they exit the target cohort

10

Incidence rates

Washout all history, no repetitive events
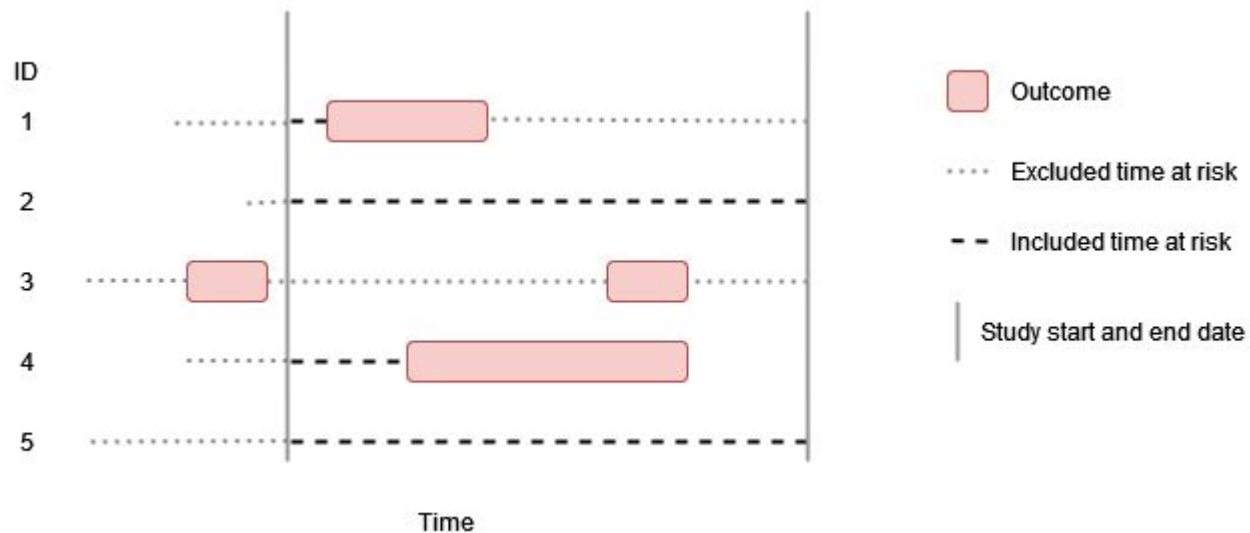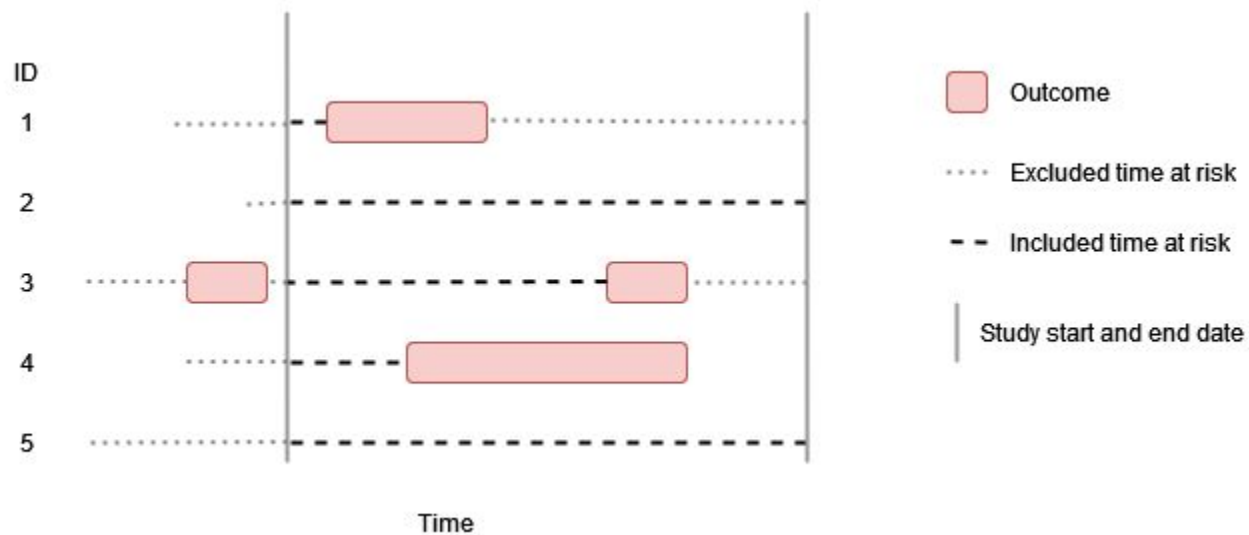
No washout, no repetitive events

No washout, no repetitive events
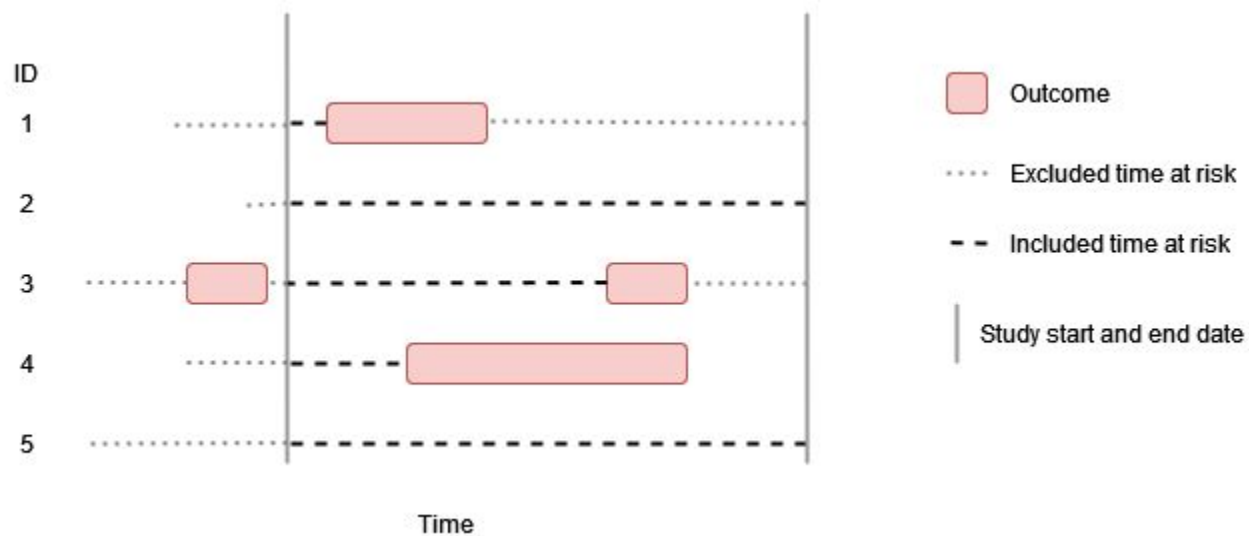
Incidence rates

Some washout, no repetitive events

Some washout, repetitive events
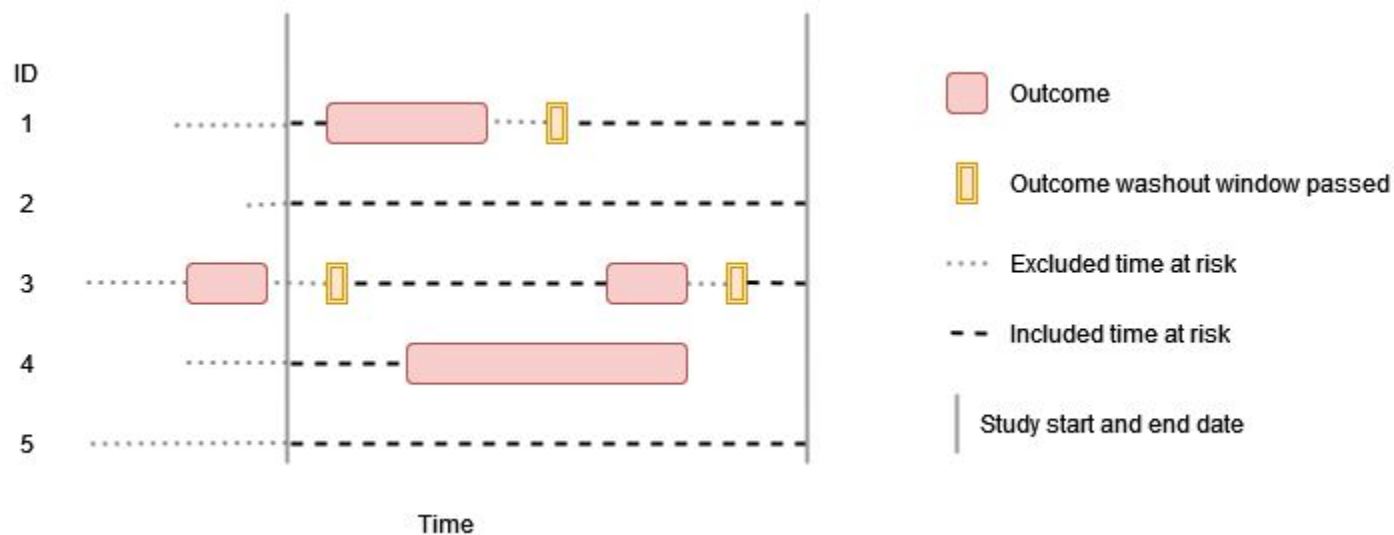
Point prevalence
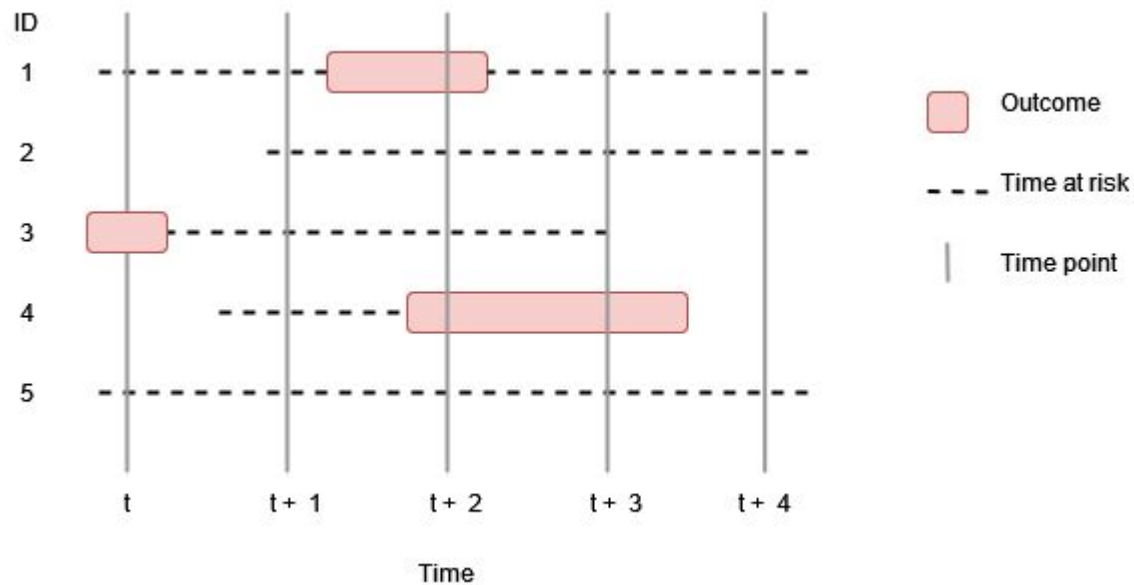
Prevalence

Period prevalence
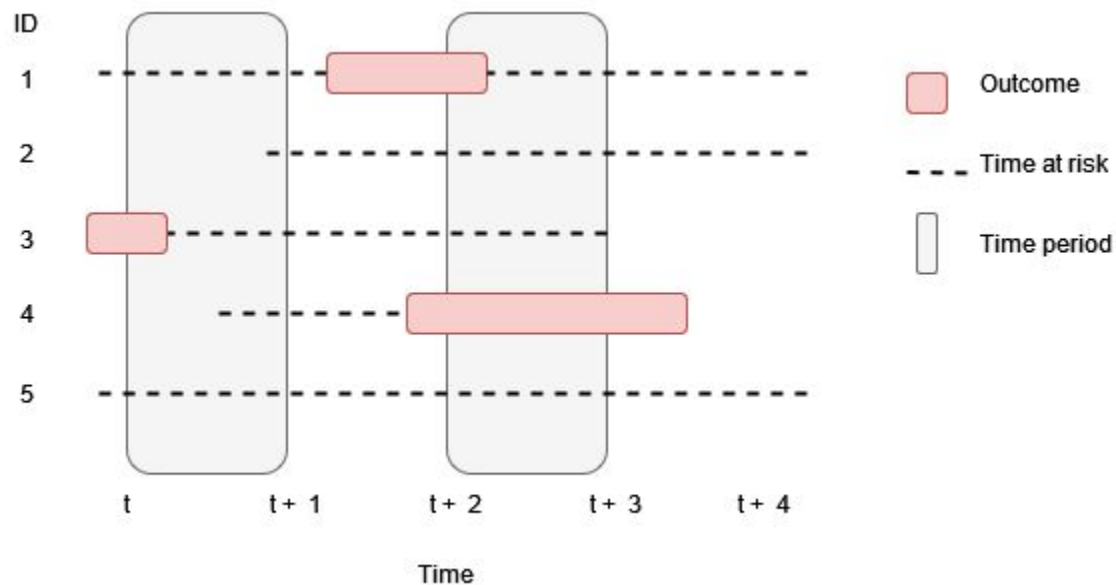
Contents

1. Concepts

2. **Interface**

3. Package Development

    a. Validation studies

    b. Benchmarking

Interface

Schema "main" ← **cdm** → Schema "Results"
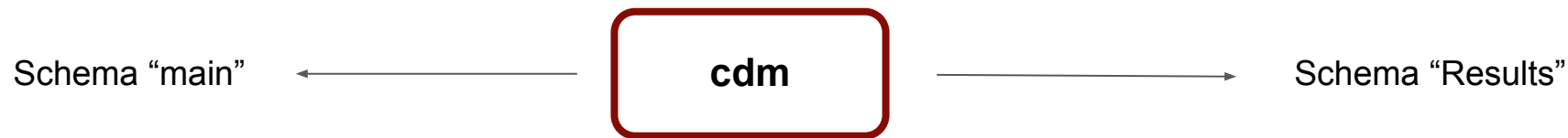
Objeto de referencia con todas
las tablas de nuestra BD

generateDenominatorCohortSet()

> cdm <- generateDenominatorCohortSet(cdm,
+   name= "dpop")

Genera una tabla con el
denominador en el objeto "cdm"

> cdm$dpop %>%
+   glimpse()

```
# Source:   table<main.denominator> [?? x 4]
# Database: DuckDB 0.7.1 [braventos@Windows 10 x64:R 4.1.2/:memory:]
   cohort_definition_id subject_id cohort_start_date cohort_end_date
                  <int> <chr>      <date>            <date>
1                     1 1          1958-12-18        1979-05-15
2                     1 2          1967-04-25        1991-11-08
3                     1 3          1985-10-12        2027-05-02
4                     1 4          1985-02-02        2052-08-14
5                     1 6          1960-09-25        2054-10-30
6                     1 8          1962-10-21        2019-04-25
7                     1 10         1954-02-19        2051-01-16
8                     1 11         1977-01-28        2067-02-09
9                     1 12         1964-06-15        2004-11-22
10                    1 13         1963-05-14        2028-06-20
# i more rows
# i Use `print(n = ...)` to see more rows
```

.

```
> cdm <- generateDenominatorCohortSet(cdm = cdm,
+                      name = "denominator",
+                      cohortDateRange=as.Date(c("2020-03-01","2021-12-01")),
+                      daysPriorObservation = 365
+                      )


> cdm$denominator%>%
+   glimpse()
```

```
Rows: ??
Columns: 4
Database: DuckDB 0.7.1 [braventos@Windows 10 x64:R 4.1.2/:memory:]
$ cohort_definition_id <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ subject_id           <chr> "11", "20", "51", "110", "112", "118", "125", "159", "168", "218", "220", "234", "258", "262", "27~
$ cohort_start_date    <date> 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2~
$ cohort_end_date      <date> 2021-12-01, 2021-12-01, 2021-04-07, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2~
```

```
> cdm <- generateDenominatorCohortSet(cdm = cdm,
+                       name = "denominator",
+                       cohortDateRange=as.Date(c("2020-03-01","2021-12-01")),
+                       daysPriorObservation = c(0,365),
+                       ageGroup = list(
+                         c(0,49),
+                         c(50,100)
+                       ),
+                       sex=c("Female", "Male", "Both"),
+                       requirementInteractions = TRUE
+                       )
```

```
cdm$denominator %>%
+   glimpse()
```

```
Rows: ??
Columns: 4
Database: DuckDB 0.7.1 [braventos@windows 10 x64:R 4.1.2/:memory:]
$ cohort_definition_id <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ subject_id           <chr> "11", "20", "51", "110", "112", "118", "125", "159", "168", "218", "220", "234", "258", "262", "27~
$ cohort_start_date    <date> 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2~
$ cohort_end_date      <date> 2021-12-01, 2021-12-01, 2021-04-07, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2~
```

```
> cdm <- generateDenominatorCohortSet(cdm = cdm,
+                      name = "denominator_strata",
+                      cohortDateRange=as.Date(c("2020-03-01","2021-12-01")),
+                      daysPriorObservation = c(0,365),
+                      ageGroup = list(
+                        c(0,49),
+                        c(50,100)
+                      ),
+                      sex=c("Female", "Male", "Both"),
+                      requirementInteractions = TRUE
+                      )
```

```
cdm$denominator %>%
+   glimpse()
```

```
Rows: ??
Columns: 4
Database: DuckDB 0.7.1 [braventos@windows 10 x64:R 4.1.2/:memory:]
$ cohort_definition_id <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ subject_id           <chr> "11", "20", "51", "110", "112", "118", "125", "159", "168", "218", "220", "234", "258", "262", "27~
$ cohort_start_date    <date> 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2~
$ cohort_end_date      <date> 2021-12-01, 2021-12-01, 2021-04-07, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2~
```

generateTargetDenominatorCohortSet()

```
> cdm <- generateTargetDenominatorCohortSet(cdm,
+                         name= "denominator_target",
+                         targetCohortTable = "target",
+                         targetCohortId = 1
+ )

cdm$denominator_target %>%
+   glimpse()
```

Cohorte "target" previamente instanciada en el objeto "cdm"

```
Rows: ??
Columns: 4
Database: DuckDB 0.7.1 [braventos@windows 10 x64:R 4.1.2/:memory:]
$ cohort_definition_id <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ subject_id           <chr> "11", "20", "51", "110", "112", "118", "125", "159", "168", "218", "220", "234", "258", "262", "27~
$ cohort_start_date    <date> 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2020-03-01, 2~
$ cohort_end_date      <date> 2021-12-01, 2021-12-01, 2021-04-07, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2021-12-01, 2~
```

Target cohort   Outcome cohort
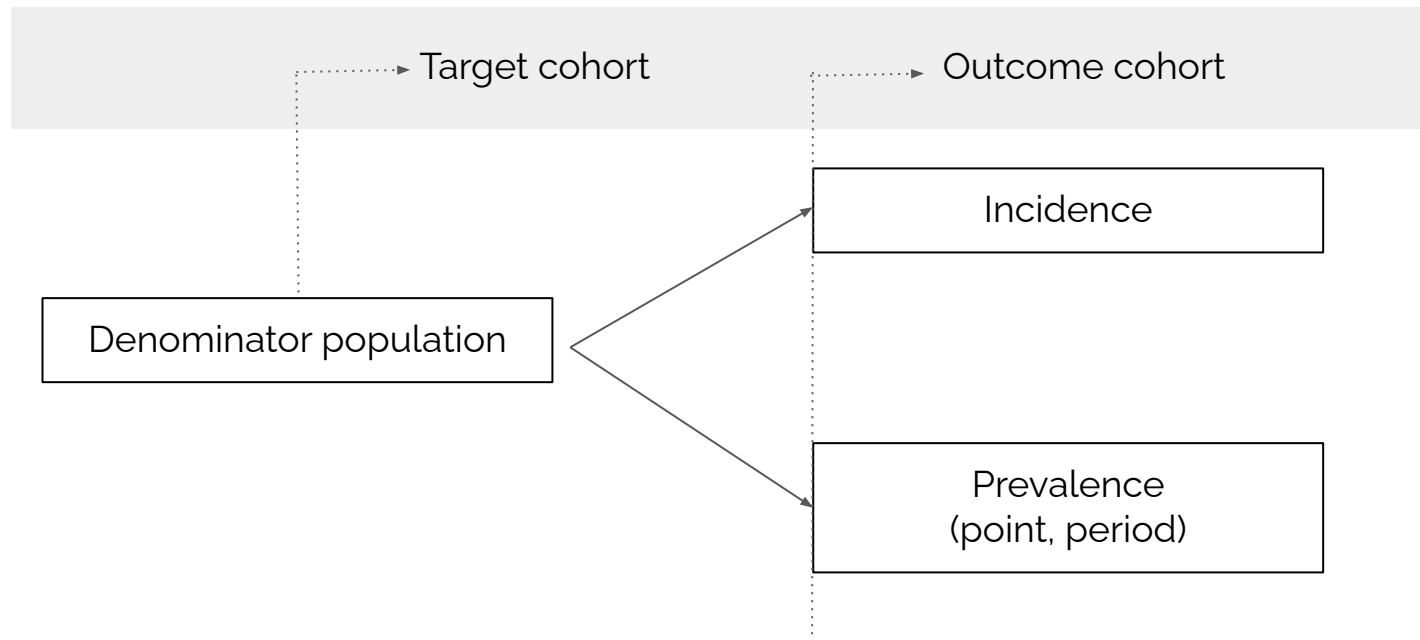
Previously instantiated
to the cdm object

Denominator population

Incidence

Prevalence
(point, period)

⚠ Target/outcome cohorts should
be carefully defined

```
> inc <- estimateIncidence(cdm = cdm,
+                          denominatorTable = "denominator",
+                          outcomeTable = "outcome",
+                          interval = "years",
+                          outcomeWashout = c(Inf, 365),
+                          repeatedEvents = FALSE
+                          )
```

Cohortes previamente instanciadas en el objeto "cdm"

```
Rows: 32
Columns: 12
$ analysis_id                <chr> "1", "1", "1", "1", "2", "2", "2", "2"~
$ n_persons                  <int> 3226, 3102, 3055, 2964, 3724, 3284, 30~
$ person_days                <dbl> 530390, 510432, 498006, 487957, 595275~
$ n_events                   <int> 1357, 1363, 1341, 1330, 1357, 1363, 13~
$ incidence_start_date       <date> 2008-01-01, 2009-01-01, 2010-01-01, 2~
$ incidence_end_date         <date> 2008-12-31, 2009-12-31, 2010-12-31, 2~
$ person_years               <dbl> 1452.1287, 1397.4867, 1363.4661, 1335.~
$ incidence_100000_pys       <dbl> 93449.02, 97532.24, 98352.28, 99554.37~
$ incidence_100000_pys_95CI_lower <dbl> 88542.54, 92422.51, 93158.06, 94275.26~
$ incidence_100000_pys_95CI_upper <dbl> 98556.64, 102850.97, 103760.73, 105052~
$ cohort_obscured            <chr> "FALSE", "FALSE", "FALSE", "FALSE", "F~
$ result_obscured            <chr> "FALSE", "FALSE", "FALSE", "FALSE", "F~
```

```
> inc <- estimateIncidence(cdm = cdm,
+                 denominatorTable = "denominator",
+                 outcomeTable = "outcome",
+                 interval = "years",
+                 outcomeWashout = c(Inf, 365),
+                 repeatedEvents = FALSE
+                 )
```

Cohortes previamente instanciadas en el objeto "cdm"

```
Rows: 32
Columns: 12
$ analysis_id                 <chr> "1", "1", "1", "1", "2", "2", "2", "2"~
$ n_persons                   <int> 3226, 3102, 3055, 2964, 3724, 3284, 30~
$ person_days                 <dbl> 530390, 510432, 498006, 487957, 595275~
$ n_events                    <int> 1357, 1363, 1341, 1330, 1357, 1363, 13~
$ incidence_start_date        <date> 2008-01-01, 2009-01-01, 2010-01-01, 2~
$ incidence_end_date          <date> 2008-12-31, 2009-12-31, 2010-12-31, 2~
$ person_years                <dbl> 1452.1287, 1397.4867, 1363.4661, 1335.~
$ incidence_100000_pys        <dbl> 93449.02, 97532.24, 98352.28, 99554.37~
$ incidence_100000_pys_95CI_lower <dbl> 88542.54, 92422.51, 93158.06, 94275.26~
$ incidence_100000_pys_95CI_upper <dbl> 98556.64, 102850.97, 103760.73, 105052~
$ cohort_obscured             <chr> "FALSE", "FALSE", "FALSE", "FALSE", "F~
$ result_obscured             <chr> "FALSE", "FALSE", "FALSE", "FALSE", "F~
```

```
> prev_point <- estimatePointPrevalence(cdm,
+                         denominatorTable = "denominator",
+                         outcomeTable = "outcome",
+                         timePoint="middle",
+                         interval = "years"
+                         )
```

Cohortes previamente instanciadas en el objeto "cdm"

```
Rows: 20
Columns: 10
$ analysis_id          <chr> "1", "1", "1", "1", "1", "2", "2", "2", "2", "2"~
$ prevalence_start_date <date> 2008-01-01, 2009-01-01, 2010-01-01, 2011-01-01,~
$ prevalence_end_date  <date> 2008-01-01, 2009-01-01, 2010-01-01, 2011-01-01,~
$ n_cases              <int> 17, 15, 28, 18, 24, 7, 8, 12, 9, 15, 17, 22, 28,~
$ n_population         <int> 2302, 2403, 2317, 2251, 2201, 1533, 1589, 1565, ~
$ prevalence           <dbl> 0.007384883, 0.006242197, 0.012084592, 0.0079964~
$ prevalence_95CI_lower <dbl> 0.004615906, 0.003786536, 0.008374124, 0.0050641~
$ prevalence_95CI_upper <dbl> 0.011795222, 0.010273991, 0.017410256, 0.0126051~
$ cohort_obscured      <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE", "FA~
$ result_obscured      <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE", "FA~
```

estimatePeriodPrevalence()

```
> prev_period <- estimatePeriodPrevalence(cdm,
+                        denominatorTable = "denominator",
+                        outcomeTable = "outcome",
+                        interval = "years",
+                        fullContribution = FALSE
+                        )
```

Cohortes previamente instanciadas en el objeto "cdm"

```
Rows: 192
Columns: 10
$ analysis_id          <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"~
$ prevalence_start_date <date> 2008-01-01, 2008-02-01, 2008-03-01, 2008-04-01,~
$ prevalence_end_date  <date> 2008-01-31, 2008-02-29, 2008-03-31, 2008-04-30,~
$ n_cases              <int> 117, 129, 151, 129, 125, 126, 154, 145, 126, 120~
$ n_population         <int> 2450, 2449, 2464, 2463, 2479, 2479, 2494, 2501, ~
$ prevalence           <dbl> 0.04775510, 0.05267456, 0.06128247, 0.05237515, ~
$ prevalence_95CI_lower <dbl> 0.03999603, 0.04450719, 0.05247785, 0.04425321, ~
$ prevalence_95CI_upper <dbl> 0.05693015, 0.06224307, 0.07145291, 0.06189121, ~
$ cohort_obscured      <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE", "FA~
$ result_obscured      <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE", "FA~
```

Common features

1. Attrition

2. Participants

3. Strata

4. Plots

Nos permite recuperar el número de personas que hemos excluido en cada paso

```
> attrition(inc)%>%
+   filter(analysis_id ==1)%>%
+   select(reason, excluded_subjects)
```

```
# A tibble: 12 x 2
   reason                                             excluded_subjects
   <chr>                                              <chr>
 1 Starting population                                <NA>
 2 Missing year of birth                              0
 3 Missing sex                                        0
 4 Cannot satisfy age criteria during the study period based ~ 0
 5 No observation time available during study period  31982
 6 Doesn't satisfy age criteria during the study period 0
 7 Prior history requirement not fulfilled during study period 0
 8 Not Male                                           9124
 9 No observation time available after applying age, prior ob~ 4430
10 Starting analysis population                       <NA>
11 Excluded due to prior event (do not pass outcome washout d~ 264
12 Not observed during the complete database interval <5
```

Strata

```
> cdm$denominator <- cdm$denominator%>%
+   PatientProfiles::addCohortIntersectFlag(
+     cdm =cdm,
+     targetCohortTable = "target",
+     targetCohortId = 1,
+     window = c(0,Inf),
+     nameStyle= "smoking"
+   )

> inc_strata <- estimateIncidence(cdm,
+                     denominatorTable = "denominator",
+                     outcomeTable = "outcome",
+                     interval = "years",
+                     strata = list("smoking")
+ )
```
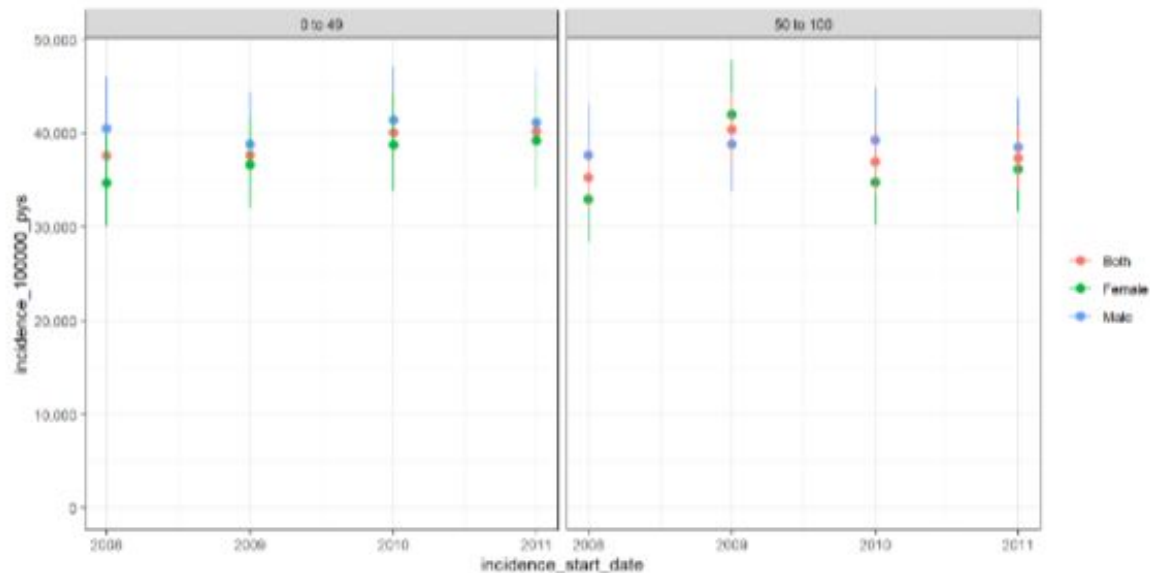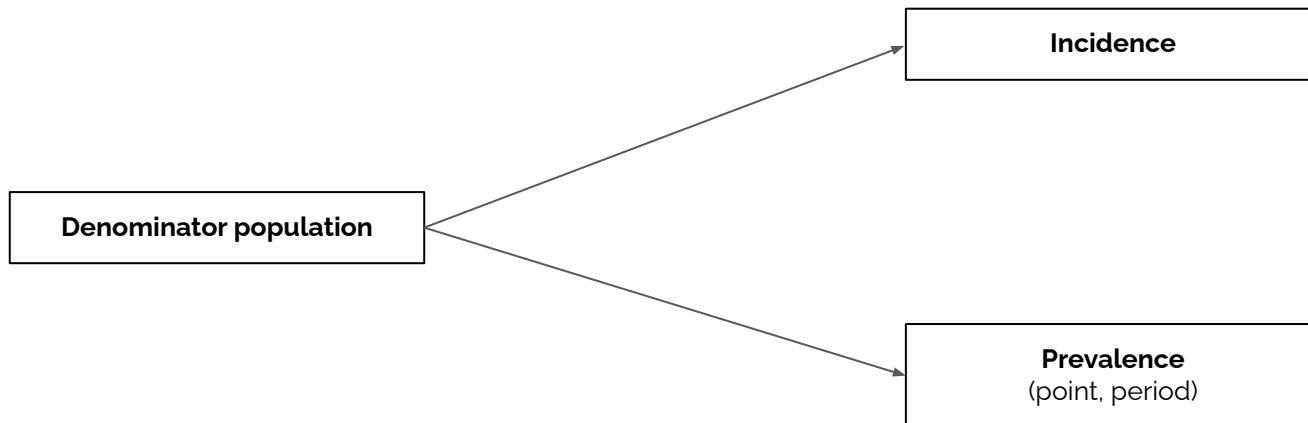
Añadimos **variable smoking** utilizando PatientProfiles

Indicamos variable por la que queremos estratificar

Plots

> plotIncidence(data,
+   facet = "denominator_age_group",
+   colour = "denominator_sex")

Función para **visualizar** los resultados

Participants

```
> inc <- estimateIncidence(cdm = cdm,
+               denominatorTable = "denominator",
+               outcomeTable = "outcome",
+               interval = "years",
+               outcomeWashout = c(Inf, 365),
+               repeatedEvents = FALSE,
+               returnParticipants = TRUE
+               )

> participants(inc, analysisId =1)
```
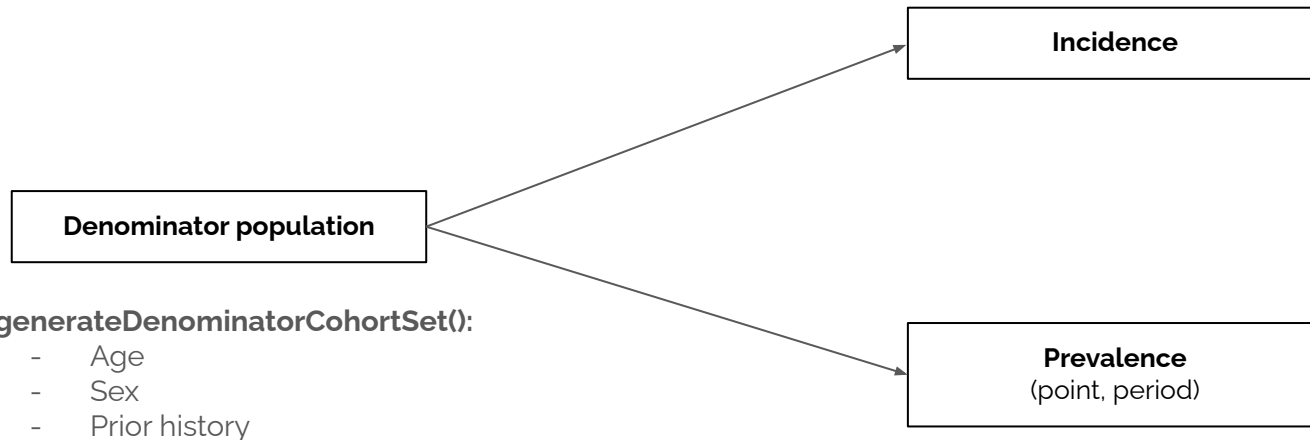
Nos permite recuperar los participantes de nuestro análisis

```
# Source:    SQL [?? x 4]
# Database: DuckDB 0.7.1 [braventos@Windows 10 x64:R 4.1.2/:memory:]
   subject_id cohort_start_date cohort_end_date outcome_start_date
   <chr>      <date>            <date>          <date>
 1 13         2008-01-01        2009-12-29      NA
 2 331        2008-01-01        2009-05-19      NA
 3 371        2010-01-20        2012-01-01      NA
 4 482        2008-01-01        2008-08-02      NA
 5 528        2011-01-10        2012-01-01      NA
```

Concepts

```
                                        ┌──────────────────────────┐
                                   ────>│        Incidence          │
                                        └──────────────────────────┘
┌──────────────────────────┐
│  Denominator population   │
└──────────────────────────┘
                                        ┌──────────────────────────┐
                                   ────>│        Prevalence         │
                                        │      (point, period)      │
                                        └──────────────────────────┘
```

Concepts

Incidence

Denominator population

Prevalence
(point, period)

generateDenominatorCohortSet():
- Age
- Sex
- Prior history

generateTarget DenominatorCohortSet():
- Based on a different cohort

estimateIncidence():
- outcomeWashout
- repetitiveEvents

Incidence

→ **plotIncidence()**

Denominator population

Prevalence
(point, period)

generateDenominatorCohortSet():
- Age
- Sex
- Prior history

generateTarget DenominatorCohortSet():
- Based on a different cohort

Concepts

estimateIncidence():
- outcomeWashout
- repetitiveEvents

```
Incidence
```
→ plotIncidence()

```
Denominator population
```

generateDenominatorCohortSet():
- Age
- Sex
- Prior history

generateTarget DenominatorCohortSet():
- Based on a different cohort

```
Prevalence
(point, period)
```
→ plotPrevalence()

estimatePointPrevalence():
- timePoint

estimatePeriodPrevalence()
- fullContribution

estimateIncidence():
- outcomeWashout
- repetitiveEvents

| **Incidence** | → **plotIncidence()**

- Attrition
- Participants
- Stratas

Denominator population

generateDenominatorCohortSet():
- Age
- Sex
- Prior history

generateTarget DenominatorCohortSet():
- Based on a different cohort

| **Prevalence** (point, period) | → **plotPrevalence()**

estimatePointPrevalence():
- timePoint

estimatePeriodPrevalence()
- fullContribution

Contents

ORIGINAL ARTICLE | 🔓 Open Access | (cc) (i) (=) ($)

IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model

Berta Raventós, Martí Català, Mike Du, Yuchen Guo, Adam Black, Ger Inberg, Xintong Li, Kim López-Güell, Danielle Newby, Maria de Ridder, Cesar Barboza, Talita Duarte-Salles ... See all authors ⌄

Berta Raventós and Martí Català should be considered as joint first-authors

Contents

1. Concepts

2. Interface

3. **Package Development**

   a. **Validation studies**

   b. Benchmarking

Validation studies

Compare IncidencePrevalence results with previously published data:

1. IR of COVID-19 in Catalonia (Spain) and England

2. IR of Adverse Events of Special Interest (AESIs)

3. PR of use of ranitidine

Compare IncidencePrevalence results with previously published data:

CPRD
n= 40 M (Aurum)
n= 15.7 M (Gold)

n= 2.6 M

SIDIAP
n= 8.3 M

Validation studies

1. **IR of COVID-19 in Catalonia and England**

IncidencePrevalence R Package results

Official government COVID-19 websites

(*) CPRD Aurum; Data from the UK was restricted to England

1. **IR of COVID-19 in Catalonia and England**
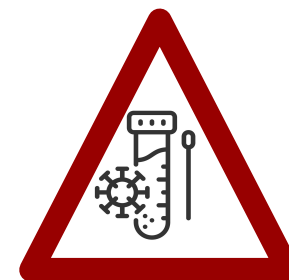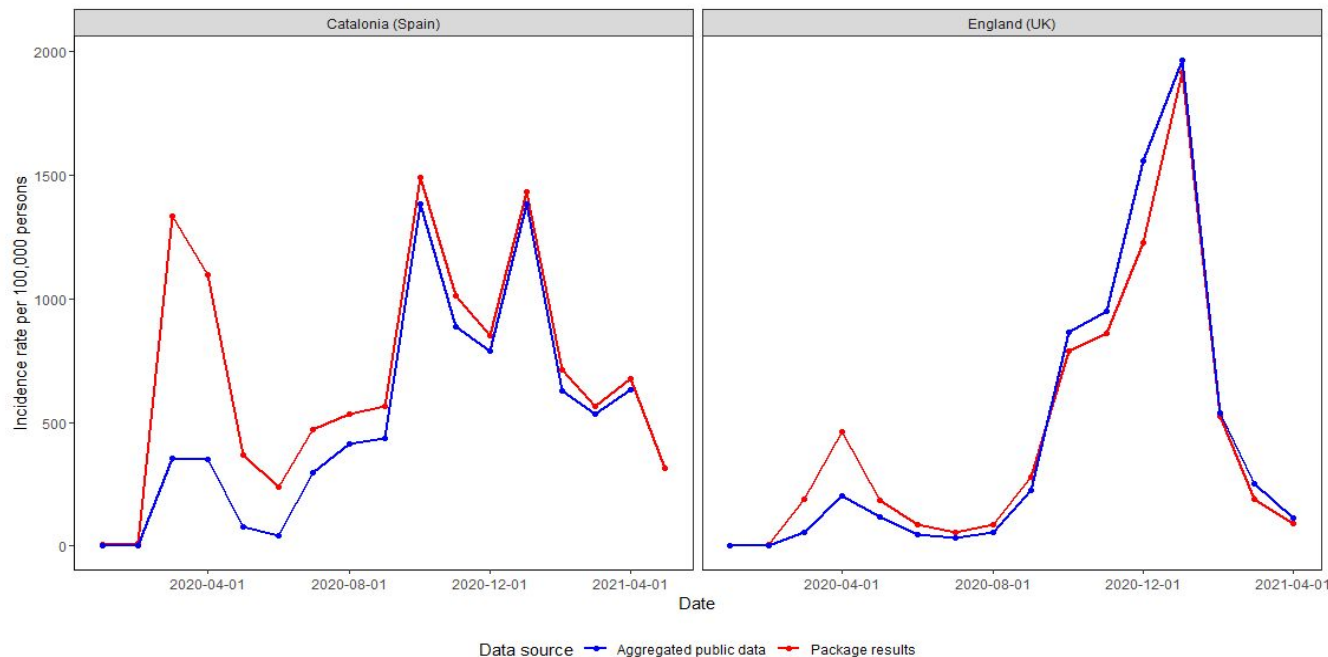
Different definitions of COVID-19

2. **IR of Adverse Events of Special Interest (AESIs)**

IncidencePrevalence R Package results

Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study
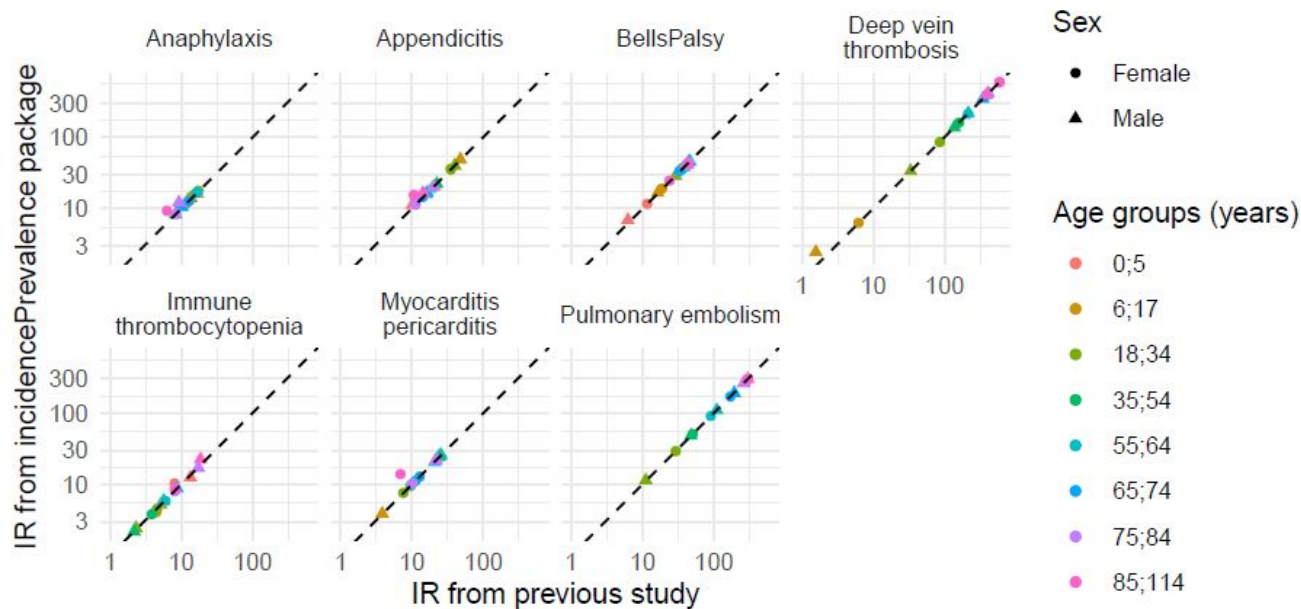
Xintong Li,[1] Anna Ostropolets,[2] Rupa Makadia,[3] Azza Shoaibi,[3] Gowtham Rao,[3] Anthony G Sena,[3,6] Eugenia Martinez-Hernandez,[4] Antonella Delmestri,[1] Katia Verhamme,[6,7] Peter R Rijnbeek,[6] Talita Duarte-Salles,[5] Marc A Suchard,[8,9] Patrick B Ryan,[2,3] George Hripcsak,[2] Daniel Prieto-Alhambra[1,6]

Previously published data

2. IR of Adverse Events of Special Interest (AESIs)
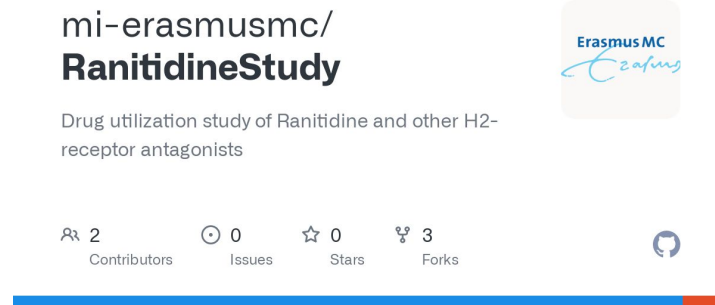
3. **PR of use of ranitidine**

IncidencePrevalence R Package results   **Previously published data**

3. **PR of use of ranitidine**
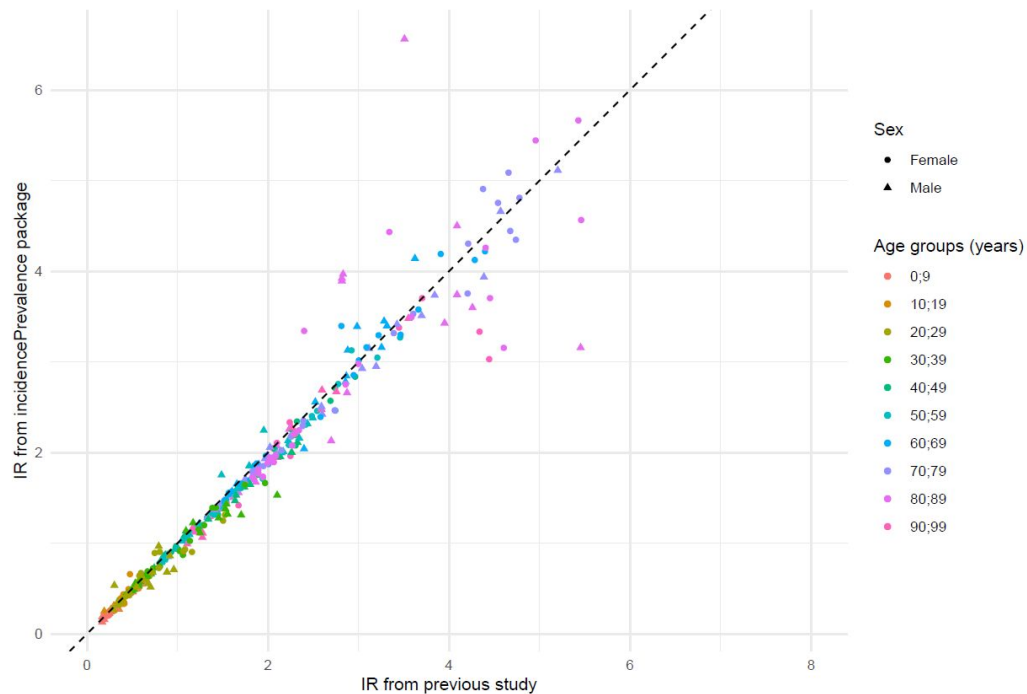
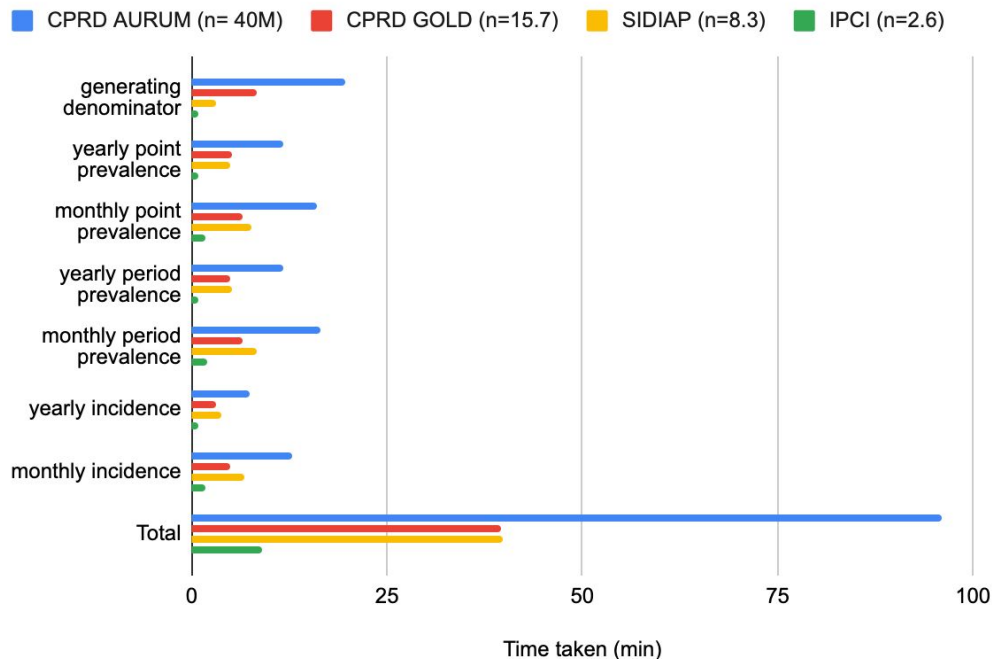# Contents

benchmarkIncidencePrevalence():

- Denominator pop.:

  - 4 age groups

  - 2 sex stratifications

- Simulated outcome (10% prev.)

Median running time:
39.7 min

⚠ Execution times rely on the size of the database and the environment

CPRD AURUM (n= 40M) CPRD GOLD (n=15.7) SIDIAP (n=8.3) IPCI (n=2.6)

generating denominator
yearly point prevalence
monthly point prevalence
yearly period prevalence
monthly period prevalence
yearly incidence
monthly incidence
Total

Time taken (min)

Thank you!