Cohort Incidence: A Software Demonstration

Christopher Knoll

Background

At OHDSI 2021, a poster was presented describing an R package that implements robust cohort incidence used in a covid-19 characterization study [1]. Since then, several releases have been made adding features to calculate rates stratified by age groups, gender, and index year.

This Software Demonstration aims to provide an introduction and technical walkthrough for designing and executing a cohort incidence characterization.

Methods

The method of cohort incidence estimation allows for multiple exposure and outcome periods per person and excludes the outcome periods (ie: immortal time) from the exposure time-at-risk to reduce bias in the incidence estimates.

To illustrate the method, the following example presents a single-patient timeline with multiple exposures and outcomes that will be used to derive the time-at-risk and cases for this person.

Step 1: Define TAR and Clean window

The initial time at risk is derived from the exposure cohort's start and end date. Figure 1 shows how time was added to the exposure cohort episodes to create the time-at-risk periods, and the clean window was added to the outcomes to produce the clean window/immortal time periods.

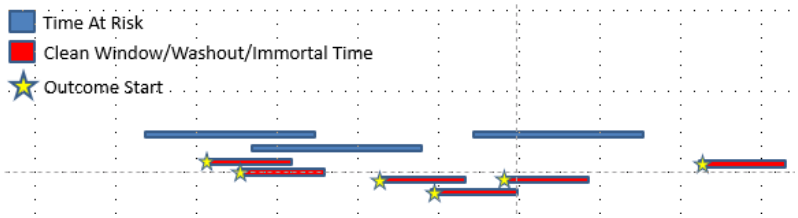

Figure 1: an example patient with multiple TAR and outcome

Step 2: Collapse TAR and Clean Windows

Any overlap that exists in the time-at-risk or clean window are collapsed so that they do not 'double count' their durations into the final incidence calculation. Figure 2 shows how the periods from Figure 1 are combined into the analysis-ready periods. Note, cases that appear during clean window periods are excluded from the calculation.

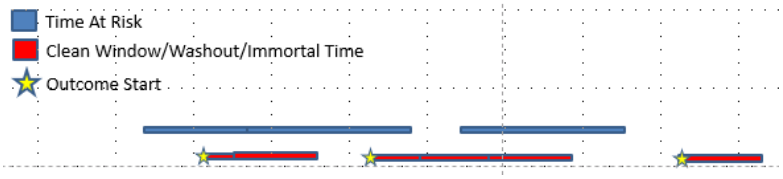

Figure 2: Merged TAR and Clean Windows

Step 3: Remove Immortal Time

If a person can not experience an outcome during the immortal time, then the immortal time should be removed from the time-at-risk. This is accomplished by subtracting the red immortal time from the blue time at risk, seen in figure 3.

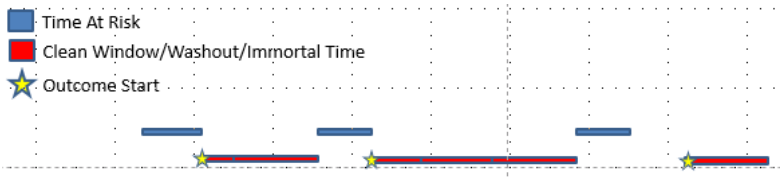

Figure 3: Exclude Immortal Time From Time-At-Risk

Step 4: Calculate Incidence

The incidence rate and proportions are calculated using the remaining time at risk and cases (shown in Figure 4):

Figure 4: Final Time-At-Risk and Cases

Incidence Rate = cases / sum(time-at-risk)

Incidence Proportion = count(distinct people w. cases) / count(distinct people w. TAR)

Note: people who have all time at risk excluded are excluded from calculation.

Results

We have produced an R package and published it to GitHub, complete with documentation and example vignettes [2]. This implementation includes R6 classes that are used to define the targets, outcomes, time-at-risk, target-outcome-tar triplets, and subgroups. These elements are assembled into a 'incidence design' which are executed on a CDM data source.

Example 1: defining targets and outcomes

For this example, we define two target populations and two outcomes of interest. Targets and outcomes reference cohort definitions, but outcomes additionally specify the clean window.

```
targets <- list(createCohortRef(id=2277, name="CPI - Any malignancy"),
                ccreateCohortRef(id=2985, name="CPI - Lung Cancer"))
outcomes <- list(createOutcomeDef(id=1,name="Skin lesion, eruption", cohortId=2284, cleanWindow=9999),
                 createOutcomeDef(id=2,name="Inflammatory dermatosis", cohortId=2285, cleanWindow=9999),
```

Example 2: define time-at-risk

The CohortIncidence package defines an R6 class to represent the time-at-risk parameters. You define a time-at-risk by specifying if the TAR should start with the target cohort's start or end date, and an offset.

The TAR end is specified in a similar way. For this example, we define 3 TARs, which all start with the target's start date and extend for 90, 180 and 365 days.

```
tars <- list(createTimeAtRiskDef(id=1, startWith="start", endWith="start", endOffset = 90),
        createTimeAtRiskDef(id=2, startWith="start", endWith="start", endOffset = 180),
        createTimeAtRiskDef(id=3, startWith="start", endWith="start", endOffset = 365)
);
```

Example 3: Using subgroups

Subgroups allow incidence calculations to be performed based on when a time-at-risk begins during a period defined by a cohort. For example, if we define a cohort to contain people with cardiovascular disorders, we can use this cohort as a subgroup to calculate incidence of people within our target cohorts that also exist in the CV cohort.

```
subgroups <- list(createCohortSubgroup(id=1, name="Cardiovascular Disorder", cohortRef =
createCohortRef(id=41)))
```

Example 4: Defining analysis T-O-TAR triplets

Once the targets, outcomes, and TARs have been defined, the combinations to analyze are specified in a 'IncidenceAnalysis' class. An incidence design can specify one or more analysis groups to control which targets should be used with which outcome and TAR, but the normal practice is to calculate all T-O-TAR pairs:

```
analysis1 <- createIncidenceAnalysis(targets = sapply(targets, function(t) { return(t$id); }),
                                     outcomes = sapply(outcomes, function(o) { return(o$id);}),
                                     tars = sapply(tars, function(t) { return(t$id); }));
```

Example 5: Strata Settings

Strata settings are used to specify how incidence is calculated on different strata of the target cohorts. The TARs are stratified by age, gender, and index year, and then aggregated into incidence calculation for each member of the strata.

```
strataSettings <- createStrataSettings(byGender=T, byAge=T, ageBreaks = c(17,34,65))
```

This code demonstrates that the incidence will be stratified by gender and age, using age breaks at 17, 34 and 65 years old.

Example 6: Building the Design and Executing

With all the elements defined, we then assemble the incidence design and execute it on a CDM:

```
irDesign <- createIncidenceDesign(targetDefs = targets,
                                  outcomeDefs = outcomes,
                                  tars=tars,
                                  subgroups = subgroups,
                                  analysisList = list(analysis1),
                                  strataSettings = strataSettings)


executeResults <- executeAnalysis(connectionDetails = connectionDetailsCDM,
                                  incidenceDesign = irDesign,
```

buildOptions = buildOptions)

Execution produces the following result:

```
> head(executeResults)
  REF_ID SOURCE_NAME TARGET_COHORT_DEFINITION_ID          TARGET_NAME TAR_ID TAR_START_WITH
1      1    optumDOD                       2987 CPI - Bladder cancer      2          start
2      1    optumDOD                       2987 CPI - Bladder cancer      2          start
3      1    optumDOD                       2987 CPI - Bladder cancer      2          start
4      1    optumDOD                       2987 CPI - Bladder cancer      2          start
5      1    optumDOD                       2987 CPI - Bladder cancer      3          start
6      1    optumDOD                       2987 CPI - Bladder cancer      3          start
  TAR_START_OFFSET TAR_END_WITH TAR_END_OFFSET SUBGROUP_ID SUBGROUP_NAME OUTCOME_ID
1                0        start            180           0           All          3
2                0        start            180           0           All          3
3                0        start            180           0           All          3
4                0        start            180           0           All          3
5                0        start            365           0           All          3
6                0        start            365           0           All          3
  OUTCOME_COHORT_DEFINITION_ID                OUTCOME_NAME CLEAN_WINDOW AGE_ID AGE_GROUP_NAME
1                         2388 Autoimmune bullous dermatosis          9999      2        17 - 33
2                         2388 Autoimmune bullous dermatosis          9999      3        34 - 64
3                         2388 Autoimmune bullous dermatosis          9999      4           >=65
4                         2388 Autoimmune bullous dermatosis          9999     NA           <NA>
5                         2388 Autoimmune bullous dermatosis          9999      2        17 - 33
6                         2388 Autoimmune bullous dermatosis          9999      3        34 - 64
  GENDER_ID GENDER_NAME START_YEAR PERSONS_AT_RISK_PE PERSONS_AT_RISK PERSON_DAYS_PE PERSON_DAYS
1      8507        MALE         NA                 3               3            256         256
2      8507        MALE         NA               303             302          40921       40770
3      8507        MALE         NA              2079            2071         285216      283827
4      8507        MALE         NA              2385            2376         326393      324853
5      8507        MALE         NA                 3               3            263         263
6      8507        MALE         NA               303             302          62089       61938
  PERSON_OUTCOMES_PE PERSON_OUTCOMES OUTCOMES_PE OUTCOMES INCIDENCE_PROPORTION_P100P
1                  0                0          0        0                 0.0000000
2                  0                0          0        0                 0.0000000
3                  5                3          6        3                 0.1448576
4                  5                3          6        3                 0.1262626
5                  0                0          0        0                 0.0000000
6                  0                0          0        0                 0.0000000
  INCIDENCE_RATE_P100PY
1             0.0000000
2             0.0000000
3             0.3860626
4             0.3373064
5             0.0000000
6             0.0000000
>
```

Figure 5: CohortIncidence Results

Conclusion

This submission describes the method and R package implementation for calculating robust cohort incidence. We've shown how the method excludes immortal time from the time-at-risk, thereby reducing bias in the incidence estimate. By using this R package, incidence characterization can be performed in a robust and standardized way and executed across a network of CDM sources.

References/Citations

Bibliography

[1] O. A. M. R. S. A. R. G. S. A. G. e. a. Li X, "Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study," *BMJ*, vol. 373,

2021.
[2] "CohortIncidence," [Online]. Available: https://ohdsi.github.io/CohortIncidence/.