Cohort Interval Algebra

Gowtham Rao

2022-10-05

Contents

1 Cohort

A cohort is a set of persons who satisfy one or more inclusion criteria (a phenotype) for a duration of time - and this is identified by cohort start date (start) and cohort end date (end). In OHDSI a) One person may belong to multiple cohorts, b) One person may belong to the same cohort for multiple different time periods, c) One person may not belong to the same cohort multiple times during the same period of time, d) A cohort may have zero or more members.

2 Multiple cohorts

Thus a person may belong to different cohorts, and these cohorts may or may not have shared dates. Studying population level temporal relationships between persons start/end of one cohort and start/end of another cohort may offer us useful characterization insights.

2.1 Use Cases

An example is a cohort of persons dispensed drug Hydrochlorthiazide (HCTZ), and a cohort of persons with diagnosis of Hypertension (HTN). Since most agree that HCTZ is one of first line drugs used in the management of Hypertension - we expect that - persons with the diagnosis of hypertension and the dispensation of HCTZ to be present in both cohorts. - in general, we expect the start of the HCTZ cohort to be on or after the diagnosis of Hypertension (because we expect the management to start after diagnosis), i.e. hctz.cohort_start_date >= hypertension.cohort_start_date. If we find that a large proportion of persons had HCTZ start prior to HTN diagnosis - then this may indicate index date misclassification.

Another use case maybe to study the relationship between end of a treatment and start of a new condition. e.g. if a person is on a long term treatment with HCTZ, stops the drug (and no other treatment is initiated) - we may expect to observe their hypertension to be uncontrolled after stopping HCTZ. i.e. HCTZ.cohort_end_date < uncontrolledHTN.cohort_start_date

3 Formal representation

One formal way of representing temporal relationships between multiple cohort is to use Allen's interval algebra (a calculus for temporal reasoning that was introduced by James F. Allen in 1983).

3.1 Thirteen basic relationship

There are 13 basic relationships. They are

1) **precedes (p)**

Comparator cohort *precedes* target cohort

Field in cohortRelationship table: ce_before_ts

Logic: cs < ts

Analysis id: -1

```
   |---C----|

                |----T----|
```

2) **meets (m)**

Comparator cohort *meets* target cohort

Field in cohortRelationship table: ce_on_ts

Logic: ce = ts

Analysis id: -2

```
   |---C---|

           |---T---|
```

3) **overlaps (o)**

Comparator cohort *overlaps* target cohort

Field in cohortRelationship table: ce_in_t_cs_bf_ts

Logic: ce > ts & ce < te & cs < ts

Analysis id: -3

```
      |---C---|
          |---T---|
```

4) **finished by (F)**

Comparator cohort *finished by* target cohort

Field in cohortRelationship table: ce_on_te_cs_bf_ts

Logic: ce = te & cs < ts

Analysis id: -4

```
    |-----C-----|
        |---T---|
```

5) **contains (D)**

Comparator cohort *contains* target cohort

Field in cohortRelationship table: cs_bf_ts_ce_gt_te

Logic: cs < ts & ce > te

Analysis id: -5

```
    |---C-------|
      |--T----|
```

6) **starts (s)**

Comparator cohort *starts* target cohort

Field in cohortRelationship table: cs_on_ts_ce_bf_te

Logic: cs = ts & ce < te

Analysis id: -6

```
    |---C---|
    |------T-------|
```

7) **equals (e)**

Comparator cohort *equals* target cohort

Field in cohortRelationship table: c_equals_t

Logic: cs = ts & ce = te

Analysis id: -7

```
      |---C-----|

      |---T-----|
```

8) **started by (S)**

Comparator cohort *started by* target cohort

Field in cohortRelationship table: cs_on_ts_ce_gt_te

Logic: cs = ts & ce > te

Analysis id: -8

```
    |------C-------|

    |----T-----|
```

9) **during (d)**

Comparator cohort *during* target cohort

Field in cohortRelationship table: c_in_t

Logic: cs > ts & ce < te

Analysis id: -9

```
      |---C----|
    |------T-------|
```

10) **finishes (f)**

Comparator cohort *finishes* target cohort

Field in cohortRelationship table: ce_on_te_cs_gt_ts

Logic: ce = te & cs > ts

Analysis id: -10

```
      |---C--------|

    |------T-------|
```

11) **overlapped by (O)**

Comparator cohort *overlapped by* target cohort

Field in cohortRelationship table: cs_in_t_te_in_c

Logic: te > cs & te < ce & cs > ts & cs < te

Analysis id: -11

```
          |------C-------|

     |----T-----|
```

12) **met by (M)**

Comparator cohort *met by* target cohort

Field in cohortRelationship table: cs_on_te

Logic: te = cs

Analysis id: -12

```
              |------C-------|

  |-----T-----|
```

13) **preceded by (P)**

Comparator cohort *preceded by* target cohort

Field in cohortRelationship table: cs_after_te

Logic: Te < Cs

Analysis id: -13

```
                   |------C-------|
   |----T-----|
```

3.2 Secondary relationships

101) **ends in (osd)**

Comparator cohort *ends in* target cohort

Field in cohortRelationship table: ce_in_t

Logic: ce > ts & ce < ts (Cs can be </=/> Ts)

Analysis id: -101

```
  |----C-----|
    |------T-------|   cOverlapsT (o)

  |----C-----|
    |------T-------|           cDuringT (d)

  |----C-----|
    |------T-------|        cStartsT (s)
```

102) **starts with start (seS)**

Comparator cohort *starts with start* target cohort

Field in cohortRelationship table: cs_on_ts

Logic: cs = ts (ce can be </=/> Te)

Analysis id: -102

```
|----C-----|
|------T-------|          cStartsT (s)

|----C-----|              cEqualsT (e)
|------T---|

|----C-----|              cIsStartedByT (S)
|------T-|
```

103) **starts in (dfO)**

Comparator cohort *starts in* target cohort

Field in cohortRelationship table: cs_in_t

Logic: cs > ts & cs < te (Ce can be </=/> Te)

Analysis id: -103

```
  |---C----|
|------T-------|          cDuringT (d)

  |---C--------|
|------T-------|          cFinishesT (f)
       |------C-------|
  |----T-----|            cIsOverlappedByT (O)
```

104) **ends with end (Fef)**

Comparator cohort *ends with end* target cohort

Field in cohortRelationship table: ce_on_te

Logic: ce = te (cs can be </=/> Ts)

Analysis id: -104

```
  |---C--------|           cIsFinishedByT (F)
    |----T-----|

  |------C-----|           cEqualsT (e)
  |------T-----|

  |---C--------|           cFinishesT (f)
|------T-------|
```

105) **starts before start (pmoFD)**

Comparator cohort *starts before start* target cohort

Field in cohortRelationship table: cs_before_ts

Logic: cs < ts (ce can be </=/> Ts/TE)

Analysis id: -105

```
|----C-----|
                |------T-------| cPrecedesT(p)

|----C-----|
           |------T-------|       cMeetsT (m)

  |----C-----|
       |------T-------|         cOverlapsT (o)

|------C-------|
  |---T--------|                 cIsFinishedByT (F)

|------C-------|
  |---T----|                     cContainsT (D)
```

106) **starts after start (dfOMP)**

Comparator cohort *starts after start* target cohort

Field in cohortRelationship table: cs_after_ts

Logic: cs > ts (ce can be </=/> TE)

Analysis id: -106

```
  |---C----|
|------T-------|   cDuringT (d)

  |---C--------|
|------T-------|   cFinishesT (f)

       |------C-------|
  |----T-----|    cIsOverlappedByT (O)

           |------C-------|
|----T-----|       CsInTTeInC (M)

                |------C-------|
|----T-----|       cPrecededByT (P)
```

107) **starts before end (pmoFDseSdfO)**

Comparator cohort *starts before end* target cohort

Field in cohortRelationship table: cs_before_te

Logic: cs < te (ce can be </=/> TE)

Analysis id: -107

```
|----C-----|
                |------T-------|   cPrecedesT (p)

|----C-----|
           |------T-------|        cMeetsT(m)

  |----C-----|
       |------T-------|            cOverlapsT(o)

|------C-------|
  |---T--------|                   cIsFinishedByT(F)

|------C-------|
  |---T----|                       cContainsT (D)

|----C-----|
|------T-------|                   cStartsT (s)

  |---C-----|
  |---T-----|                     cEqualsT (e)

|------C-------|
|----T-----|                      cIsStartedByT(S)

  |---C----|
|------T-------|                  cDuringT(d)

  |---C--------|
|------T-------|                  cFinishesT(f)

       |------C-------|
  |----T-----|                    cIsOverlappedByT(O)
```

108) **ends before end (pmoFDseSd)**

Comparator cohort *ends before end* target cohort

Field in cohortRelationship table: ce_before_te

Logic: ce < te (ce can be </=/> TE)

Analysis id: -108

```
|----C-----|
                |------T-------| cPrecedesT (p)


|----C-----|
           |------T-------| cMeetsT (m)


  |----C-----|
       |------T-------| cOverlapsT (o)
```

```
|----C-----|
|------T-------|  cStartsT (s)

  |---C----|
|------T-------|  cDuringT (d)
```

109) **ends after end (DSOMP)**

Comparator cohort *ends after end* target cohort

Field in cohortRelationship table: ce_after_te

Logic: ce > te (cs can be </=/> TS/TE)

Analysis id: -109

```
|------C-------|
  |---T----|       cContainsT (D)

|------C-------|
|----T-----|       cIsStartedByT (S)

       |------C-------|
  |----T-----|     cIsOverlappedByT (O)

           |------C-------|
|----T-----|       cIsMetByT (M)

                |------C-------|
|----T-----|       cPrecededByT (P)
```

110) **starts in inclusive (seSdfOM)**

Comparator cohort *starts in inclusive* target cohort

Field in cohortRelationship table: cs_window_t

Logic: cs >= ts & cs <= ts (ce can be </=/> TE)

Analysis id: -110

```
|----C-----|
|------T-------|   cStartsT (s)

  |---C-----|
  |---T-----|  cEqualsT (e)

|------C-------|
|----T-----|  cIsStartedByT (S)

  |---C----|
|------T-------| cDuringT (d)

  |---C--------|
|------T-------| cFinishesT (f)
```

```
      |------C-------|
  |----T-----|    cIsOverlappedByT (O)

          |------C-------|
|----T-----|      cIsMetByT (M)
```

111) **ends in inclusive (oFsedf)**

Comparator cohort *ends in inclusive* target cohort

Field in cohortRelationship table: ce_window_t

Logic: cs >= ts & cs <= ts (ce can be </=/> TE)

Analysis id: -111

```
  |----C-----|
       |------T-------|   cOverlapsT (o)

|------C-------|
  |---T--------|          cIsFinishedByT(F)

  |----C-----|
  |------T-------|        cStartsT (s)

  |---C-----|
  |---T-----|             cEqualsT (e)

  |----C-----|
 |------T-------|           cDuringT (d)

  |---C--------|
|------T-------|          cFinishesT (f)
```

112) **starts on or before start (pmoFDseS)**

Comparator cohort *starts on or before start* target cohort

Field in cohortRelationship table: cs_window_ts

Logic: cs < ts (cs can be </= ts)

Analysis id: -112

```
|----C-----|
               |------T-------|  cPrecedesT (p)

|----C-----|
           |------T-------|      CeOnTs (m)

  |----C-----|
       |------T-------|-      cOverlapsT (o)

|------C-------|
```

```
  |---T--------|   cIsFinishedByT (F)

|------C-------|
  |---T----|       cContainsT (D)

|----C-----|
|------T-------|    cStartsT (s)

  |---C-----|
  |---T-----|       cEqualsT (e)

|------C-------|
|----T-----|     cIsStartedByT (S)
```

113) **starts on or before end (pmoFDseSdfoM)**

Comparator cohort *starts on or before end* target cohort

Field in cohortRelationship table: cs_window_te

Logic: cs < te (cs can be </= te)

Analysis id: -113

```
|----C-----|
                |------T-------| cPrecedesT (p)

|----C-----|
           |------T-------|      cMeetsT (m)

  |----C-----|
       |------T-------|          cOverlapsT(o)

|------C-------|
  |---T--------|                 cIsFinishedByT (F)

|------C-------|
  |---T----|                     cContainsT (D)

|----C-----|
|------T-------|                 cStartsT (s)

  |---C-----|
  |---T-----|                   cEqualsT (e)

|------C-------|
|----T-----|                    cIsStartedByT (S)

  |---C----|
|------T-------|                 cDuringT (d)

  |---C--------|
|------T-------|                 cFinishesT (f)
```

```
       |------C-------|
  |----T-----|                    cIsOverlappedByT (O)

           |------C-------|
|----T-----|                      cIsMetByT (M)
```

114) **ends on or before end (pmoFsedf)**

Comparator cohort *ends on or before end* target cohort

Field in cohortRelationship table: ce_window_te

Logic: ce < te (ce can be </= te)

Analysis id: -114

```
|----C-----|
                |------T-------|  cPrecedesT (p)

|----C-----|
           |------T-------|  cMeetsT (m)

  |----C-----|
       |------T-------|  cOverlapsT (o)

|------C-------|
  |---T--------| cIsFinishedByT (F)

|----C-----|
|------T-------|  cStartsT (s)

  |---C-----|
  |---T-----| cEqualsT (e)

  |---C----|
|------T-------| cDuringT (d)

  |---C--------|
|------T-------|  cFinishesT (f)
```

115) **during inclusive (esdf)**

Comparator cohort *during inclusive* target cohort

Field in cohortRelationship table: c_within_t

Logic: ce >= ts & ce =< te

Analysis id: -115

```
  |---C-----|
  |---T-----|             cEqualsT (e)

|----C-----|
|------T-------|          cStartsT (s)
```

```
  |---C----|
|------T-------|          cDuringT (d)

  |---C--------|
|------T-------|          cFinishesT (f)
```

```
#> # A tibble: 28 x 6
#>    analysisId primary ontology    code  field             logic
#>         <dbl>   <dbl> <chr>       <chr> <chr>             <chr>
#>  1         -1       1 precedes    p     ce_before_ts      cs < ts
#>  2         -2       1 meets       m     ce_on_ts          ce = ts
#>  3         -3       1 overlaps    o     ce_in_t_cs_bf_ts  ce > ts & ce < te & cs < ts
#>  4         -4       1 finished by F     ce_on_te_cs_bf_ts ce = te & cs < ts
#>  5         -5       1 contains    D     cs_bf_ts_ce_gt_te cs < ts & ce > te
#>  6         -6       1 starts      s     cs_on_ts_ce_bf_te cs = ts & ce < te
#>  7         -7       1 equals      e     c_equals_t        cs = ts & ce = te
#>  8         -8       1 started by  S     cs_on_ts_ce_gt_te cs = ts & ce > te
#>  9         -9       1 during      d     c_in_t            cs > ts & ce < te
#> 10        -10       1 finishes    f     ce_on_te_cs_gt_ts ce = te & cs > ts
#> # ... with 18 more rows
```