Create a matched cohort to provide context for your large-scale characterisation

Context: The value of matched sampling for large-scale characterisation during phenotype diagnostics

Background: Large-scale characterisation aims to summarise all clinical events recorded in the database for a specific cohort in some window around their index date. This key step enables the researchers to have a comprehensive overview of the cohort and determine if it represents the target population. However, it usually results difficult to establish if there are unexpected imbalances, as the population in each database can be very specific.

We present two distinct ways to generate a comparator cohort: Using Achilles tables and CohortConstructor R package.

Results
Example: Large-scale characterisation of a lupus cohort in CPRD GOLD database (Nc = 7,381).

Methods
Large scale characterisation of our target cohort
- Assume we have a database with Np participants and a cohort with Nc individuals.
- We perform a large-scale characterisation of our cohort using the R package PatientProfiles.
- Only conditions that have an occurrence higher than 0.5% in our cohort within a time window w1 are included.

Contextualisation using a matched cohort
- Create a matched cohort with individuals that (1) are not included in the target cohort and (2) have the same year of birth and sex:
  ```r
cdm["matched_cohort"] <- CohortConstructor::matchCohorts(
    cohort = cdm$cohort, 
    name = "matched_cohort", 
    matchSex = TRUE, 
    matchYearOfBirth = TRUE, 
    ratio = 1)
```

Conclusion
- Proportions based on Achilles provide estimated counts if conditions are distributed uniformly, but otherwise can lead to spurious context.
- matchCohorts() can be used to create a comparable matched cohort based on age and sex.

See all the results in the Shiny App!