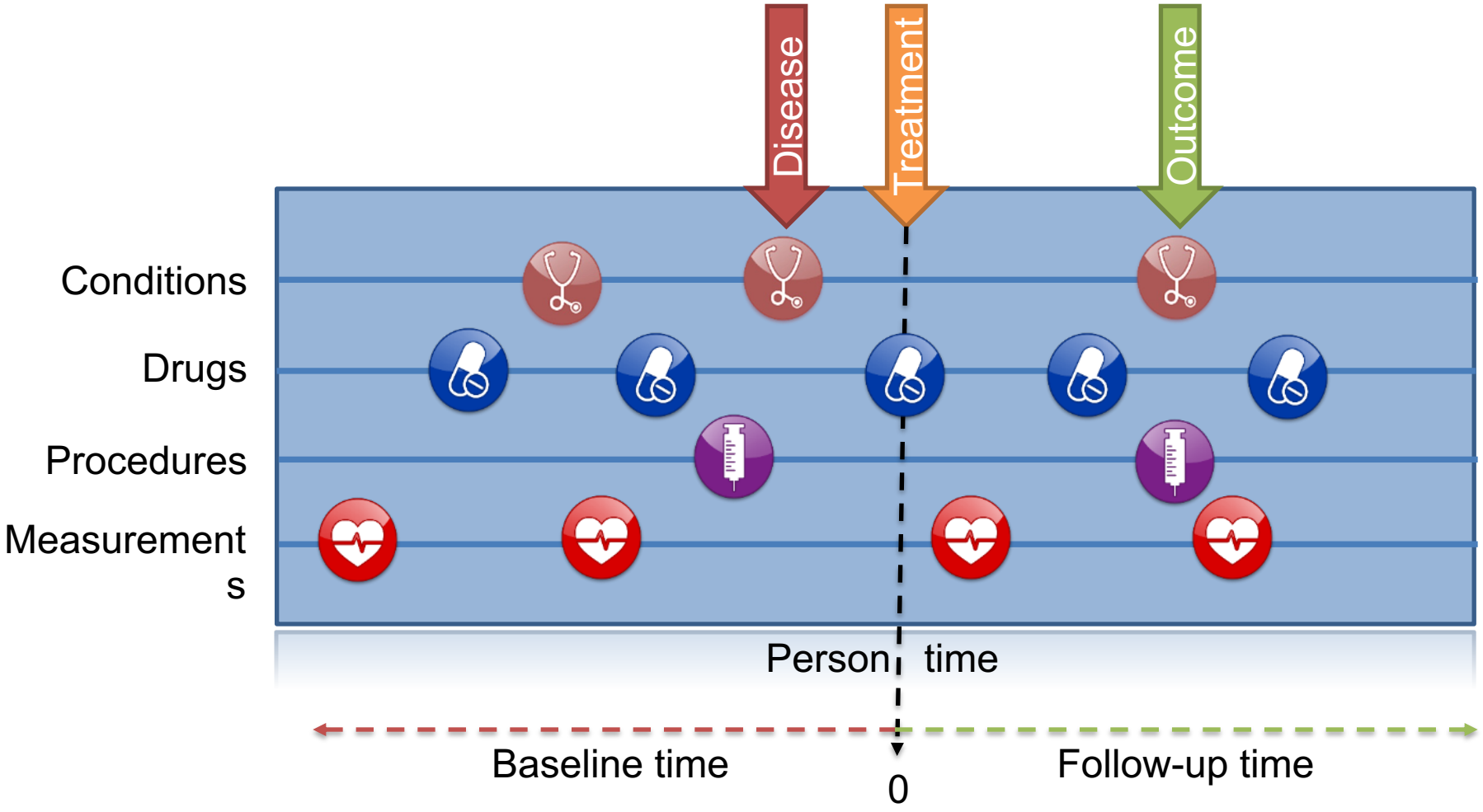
Creating cohort definitions

Asieh Golozar

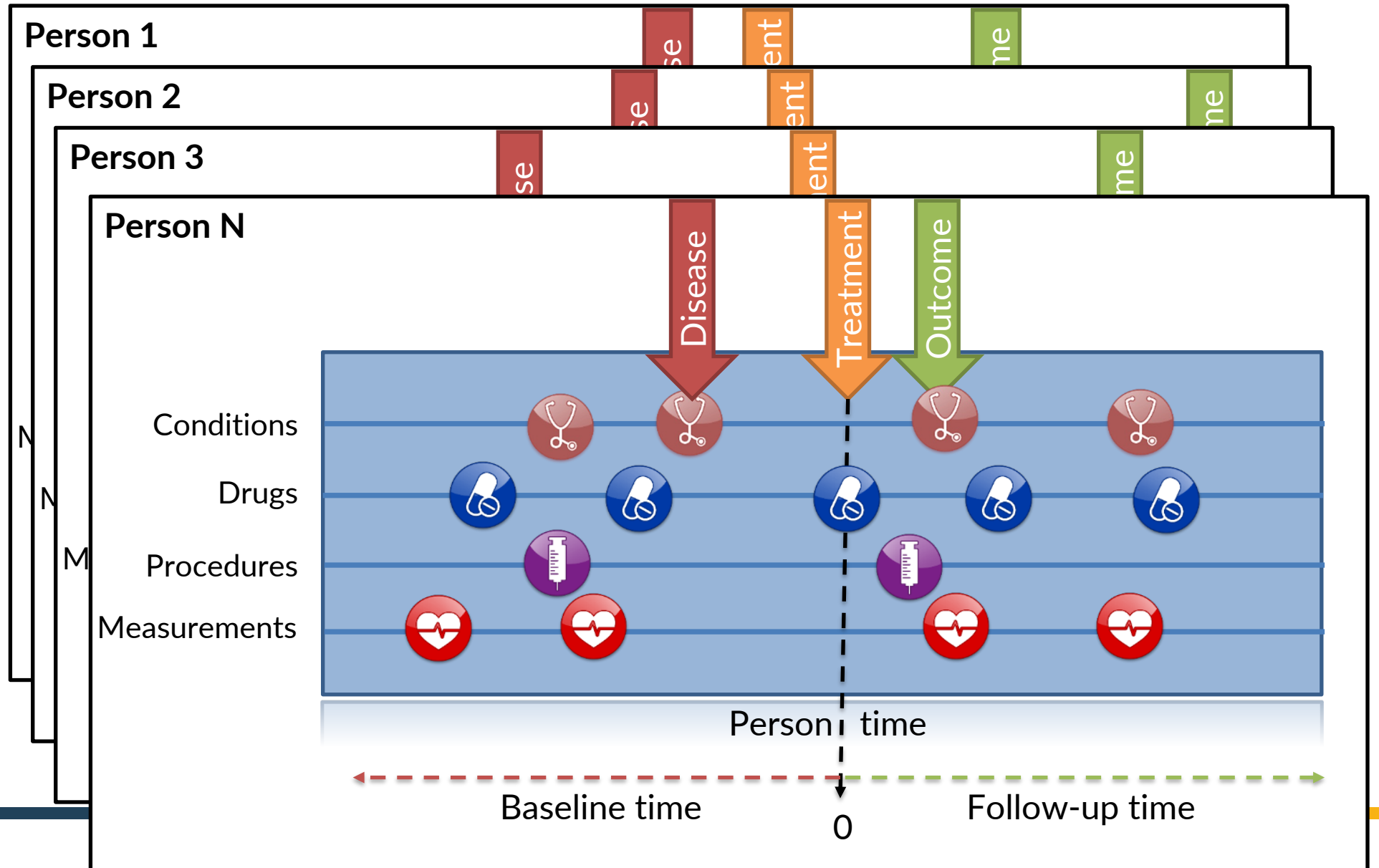
Agenda

- What is a cohort?
- Demo #1: Create cohort for ‘new users of lisinopril’ in ATLAS
- Demo #2: Create cohort for ‘acute myocardial infarction events’ in ATLAS

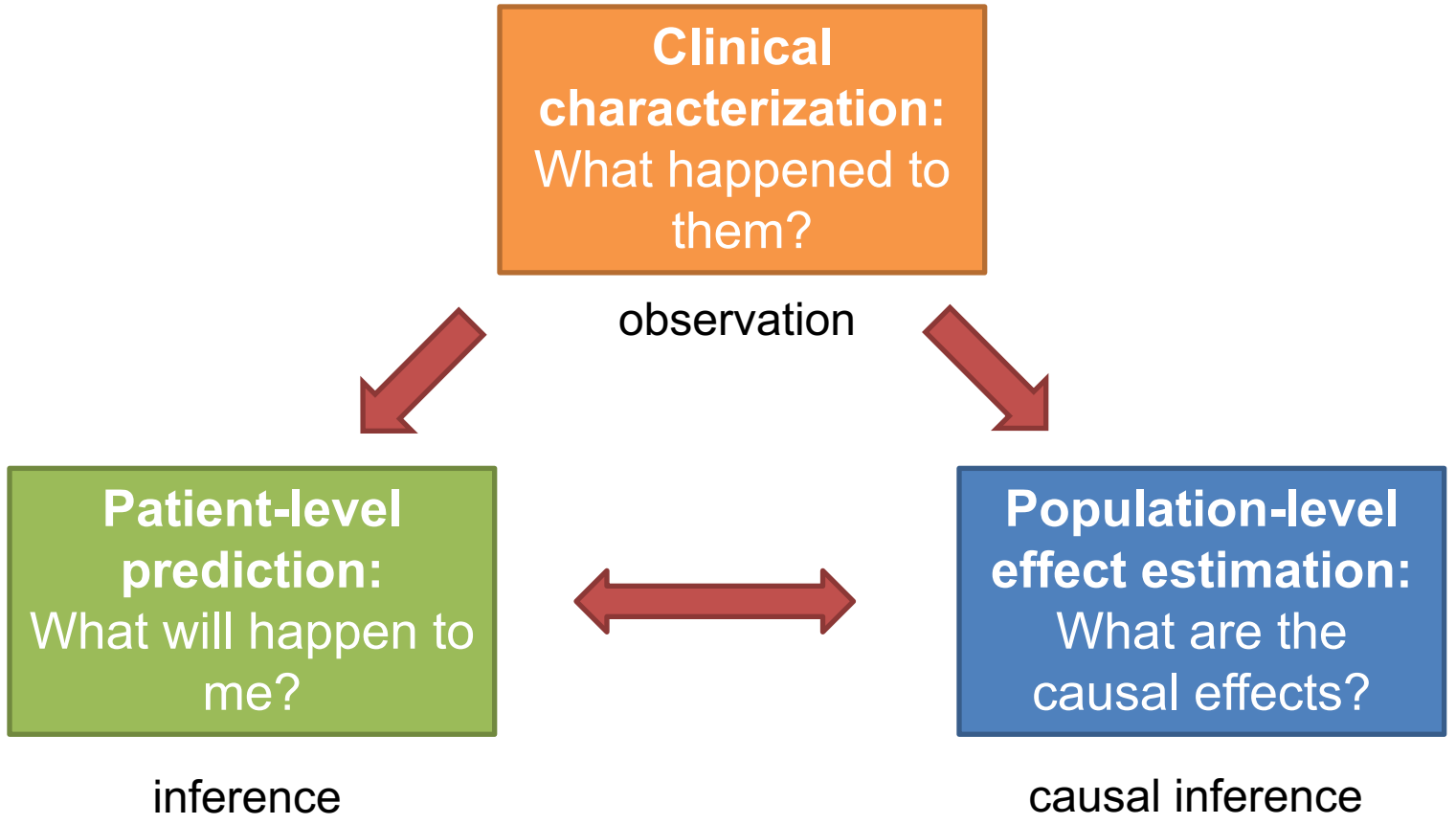
A caricature of the patient journey



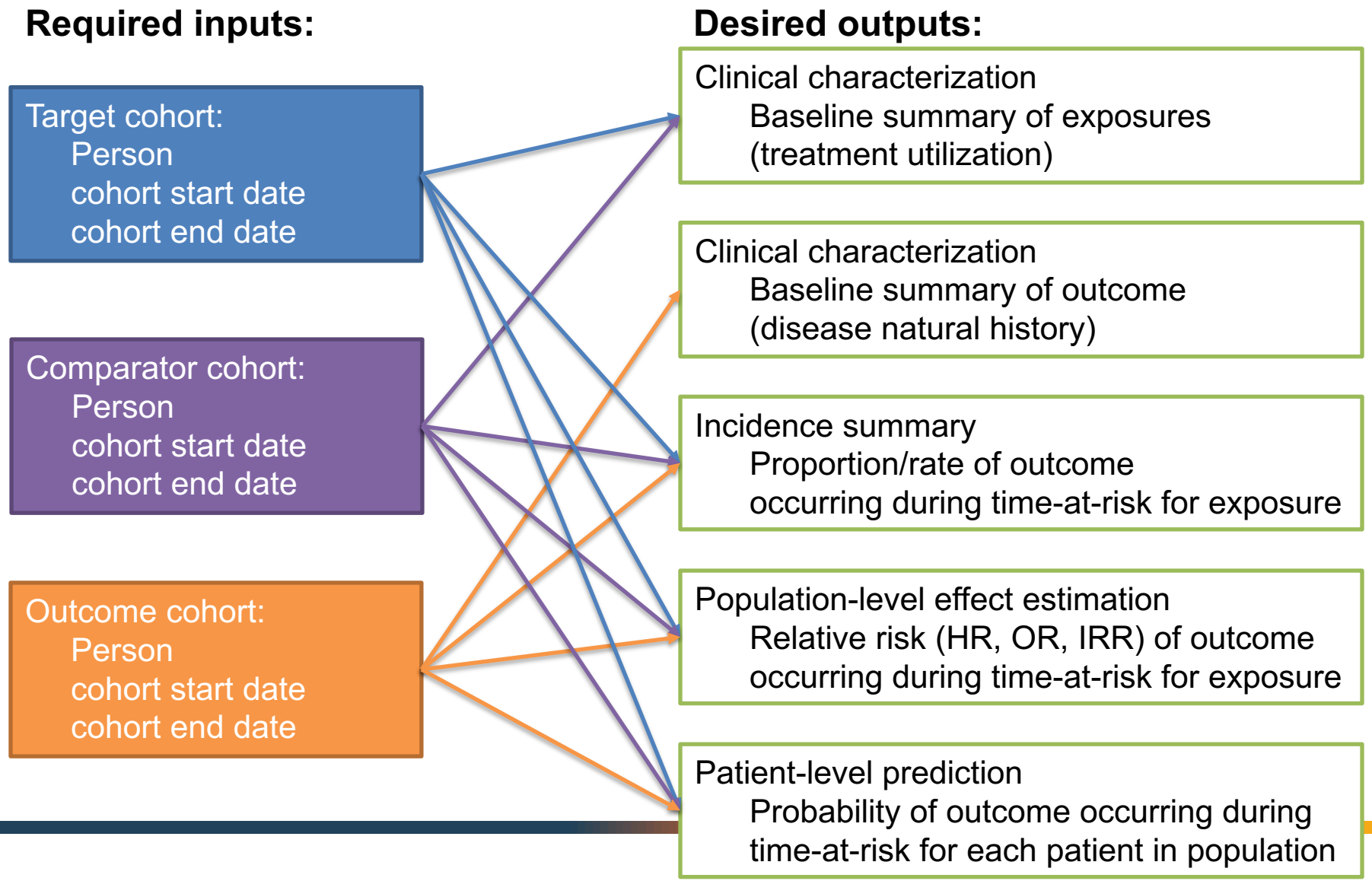
Each observational database is just an (incomplete) compilation of patient journeys



Complementary evidence to inform the patient journey



The common building block of all observational analysis: cohorts



Defining ‘phenotype’

Journal of the American Medical Informatics Association, 0(0), 2017, 1–6
doi: 10.1093/jamia/ocx110
Perspective

Perspective

High-fidelity phenotyping: richness and freedom from bias

George Hripcsak[1] and David J Albers[1]

- A phenotype is a specification of an observable, potentially changing state of an organism (as distinguished from the genotype, derived from genetic makeup).
- The term phenotype can be applied to patient characteristics inferred from electronic health record (EHR) data.
- The goal is to draw conclusions about a target concept based on raw EHR data, claims data, or other clinically relevant data.
- Phenotype algorithms – ie, algorithms that identify or characterize phenotypes – may be generated by domain exerts and knowledge engineers, or through diverse forms of machine learning to generate novel representations of data.

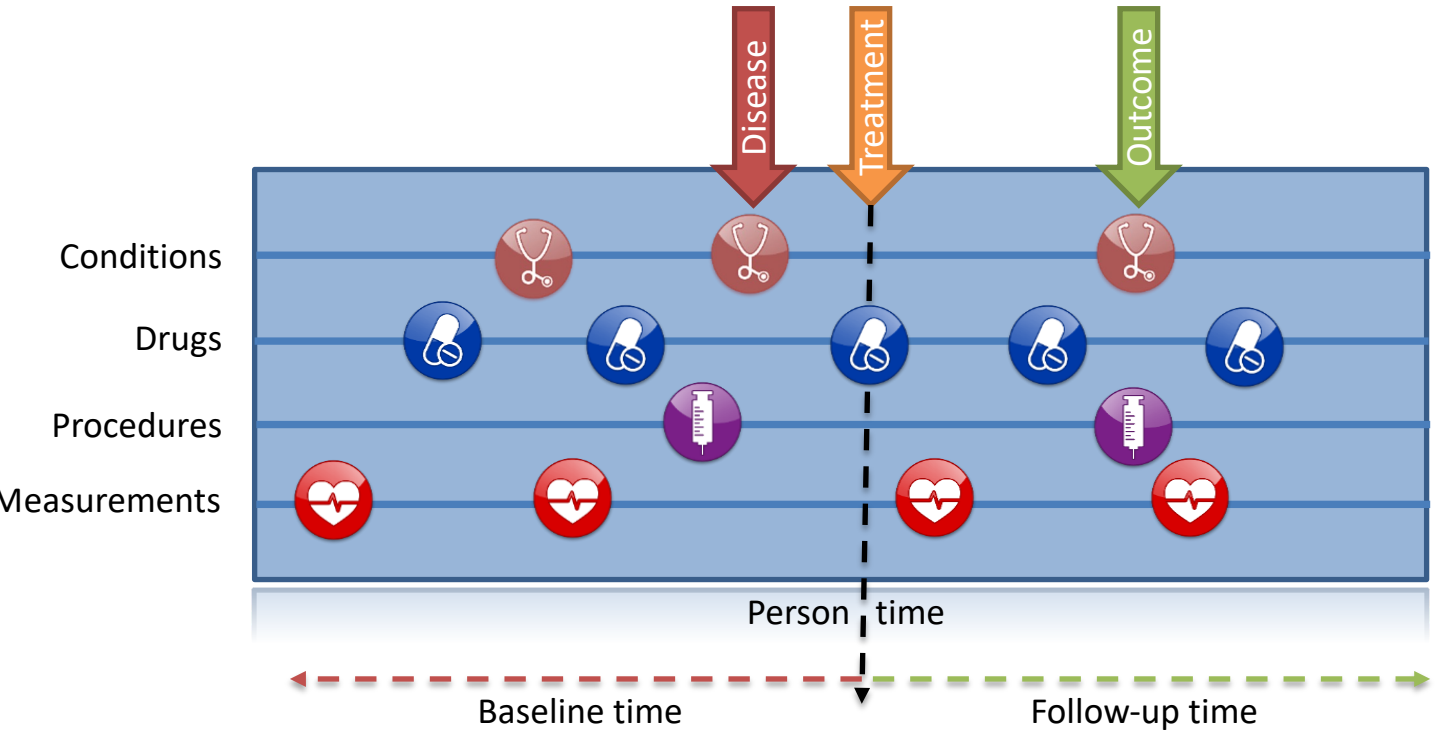
Two Approaches to Phenotyping

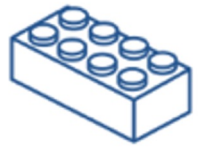
Rule-Based Phenotyping

Probabilistic Phenotyping

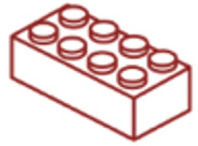
Library (Aphrodite)

Data are Like Lego Bricks for Phenotyping

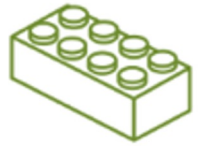




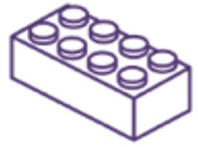
Conditions



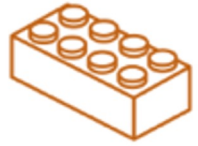
Drugs



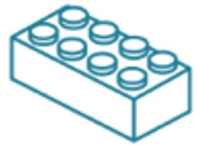
Procedures



Measurements



Observations



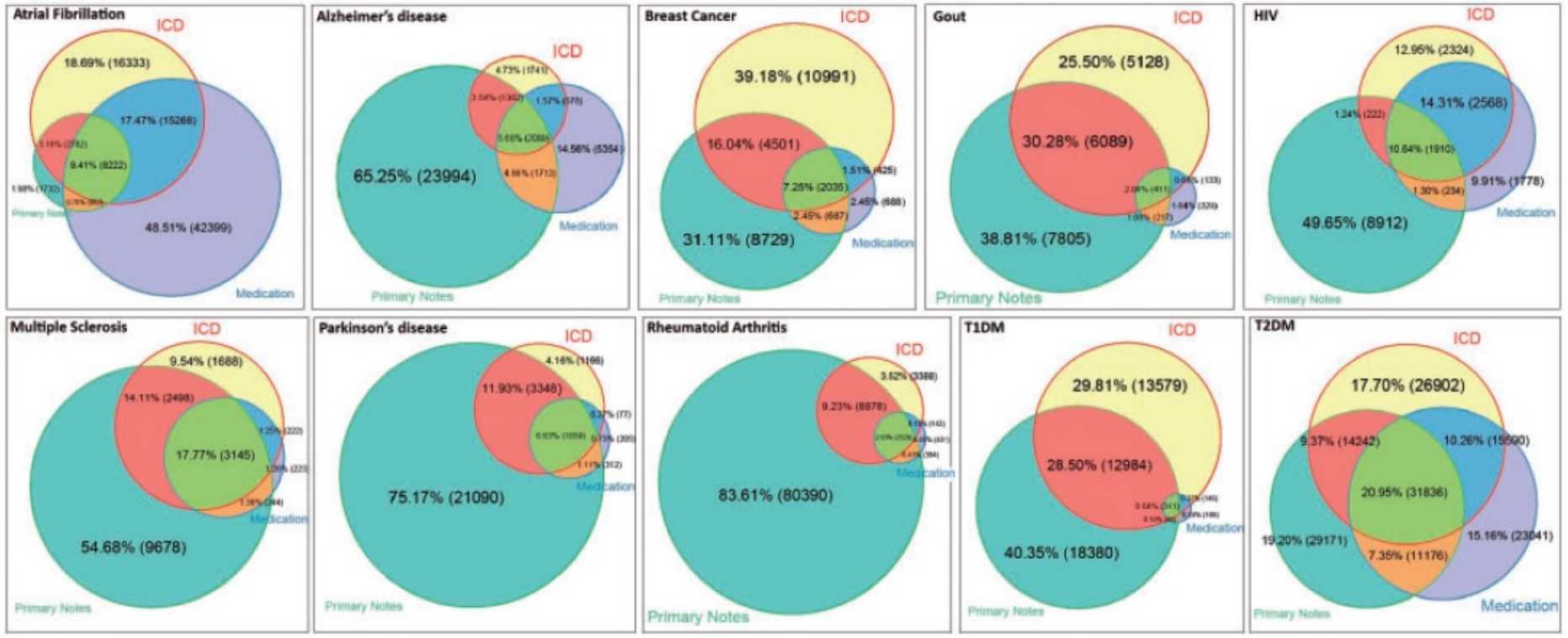
Visits

Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance

RECEIVED 8 January 2015
REVISED 14 July 2015
ACCEPTED 15 July 2015
PUBLISHED ONLINE FIRST 2 September 2015

Wei-Qi Wei[1], Pedro L Teixeira[1], Huan Mo[1], Robert M Cronin[1,2], Jeremy L Warner[1,2], Joshua C Denny[1,2]

Figure 1: Weighted Venn diagrams of the distributions of patients with ICD-9, primary notes, and specific medications. Each color represents a resource. Different area colors represent the number of patients that were found within intersecting resources.



Cohorts: The common building block of all observational analysis

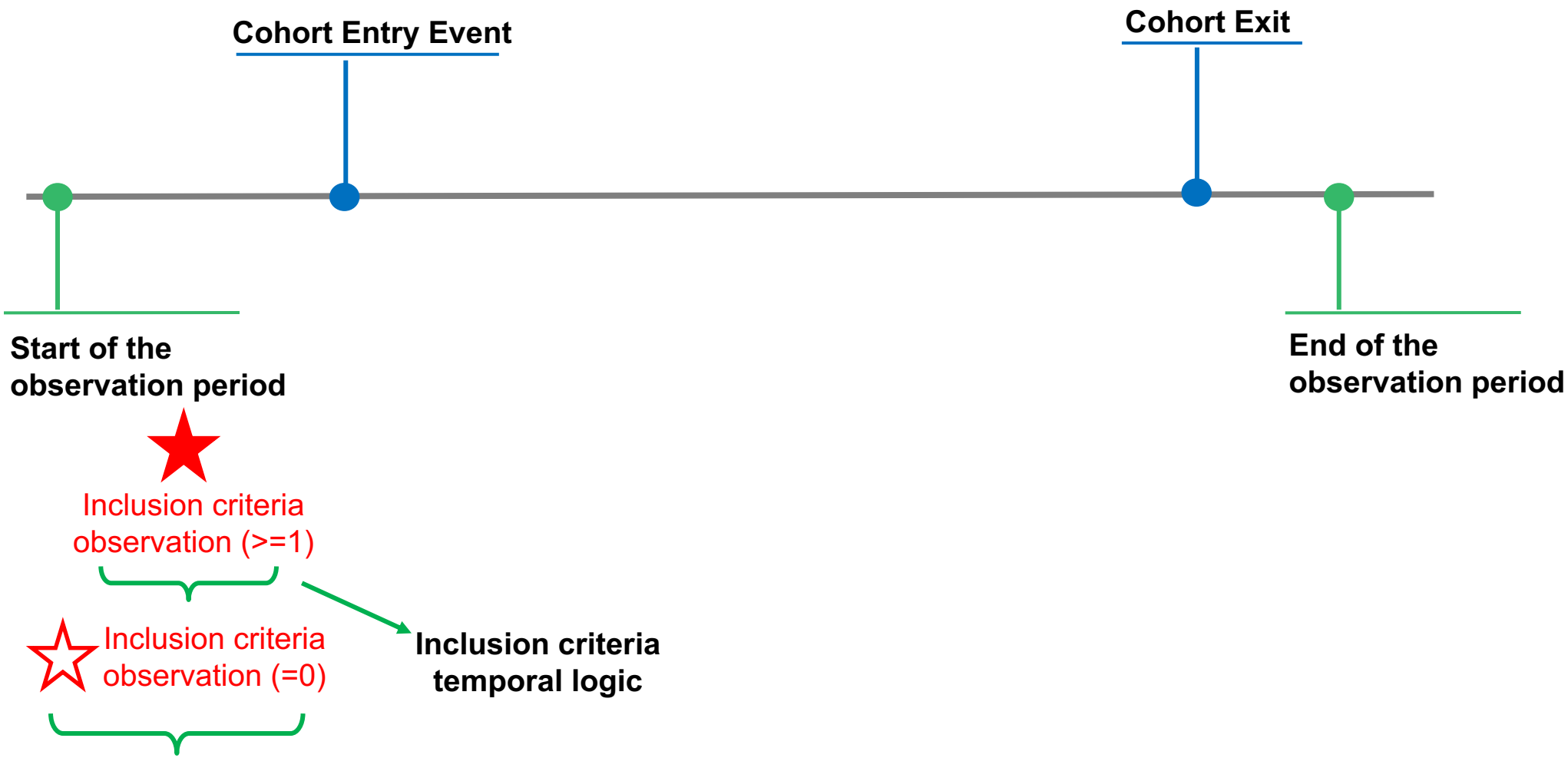
- **OHDSI's definition of 'cohort':** Cohort is a set of persons who satisfy one or more inclusion criteria for a duration of time
- **Cohort era:** a continuous period during which a person has satisfied a cohort's inclusion criteria
- **Cohort definition:** the specification for how to identify a cohort

Objective consequences based on this cohort definition:

- One person may belong to multiple cohorts
- One person may belong to the same cohort at multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- One cohort may have **zero or more** members
- A codeset is **NOT** a cohort…

…logic for how to use the codeset in a criteria is required

The Anatomy of a Cohort Definition

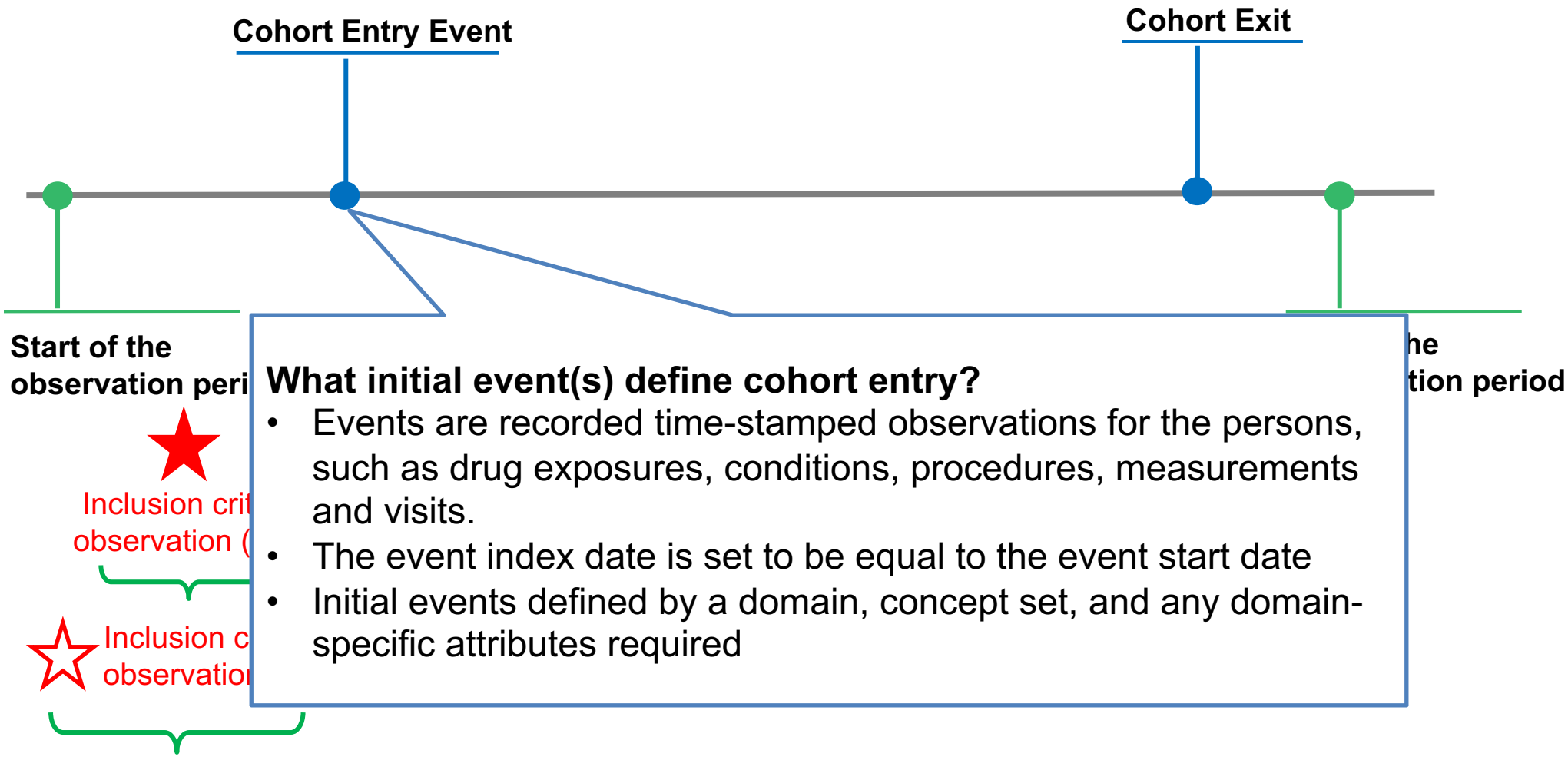


Defining Cohorts in Atlas

1 Questions to answer when defining a cohort

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?
- How should events be combined into cohort eras?

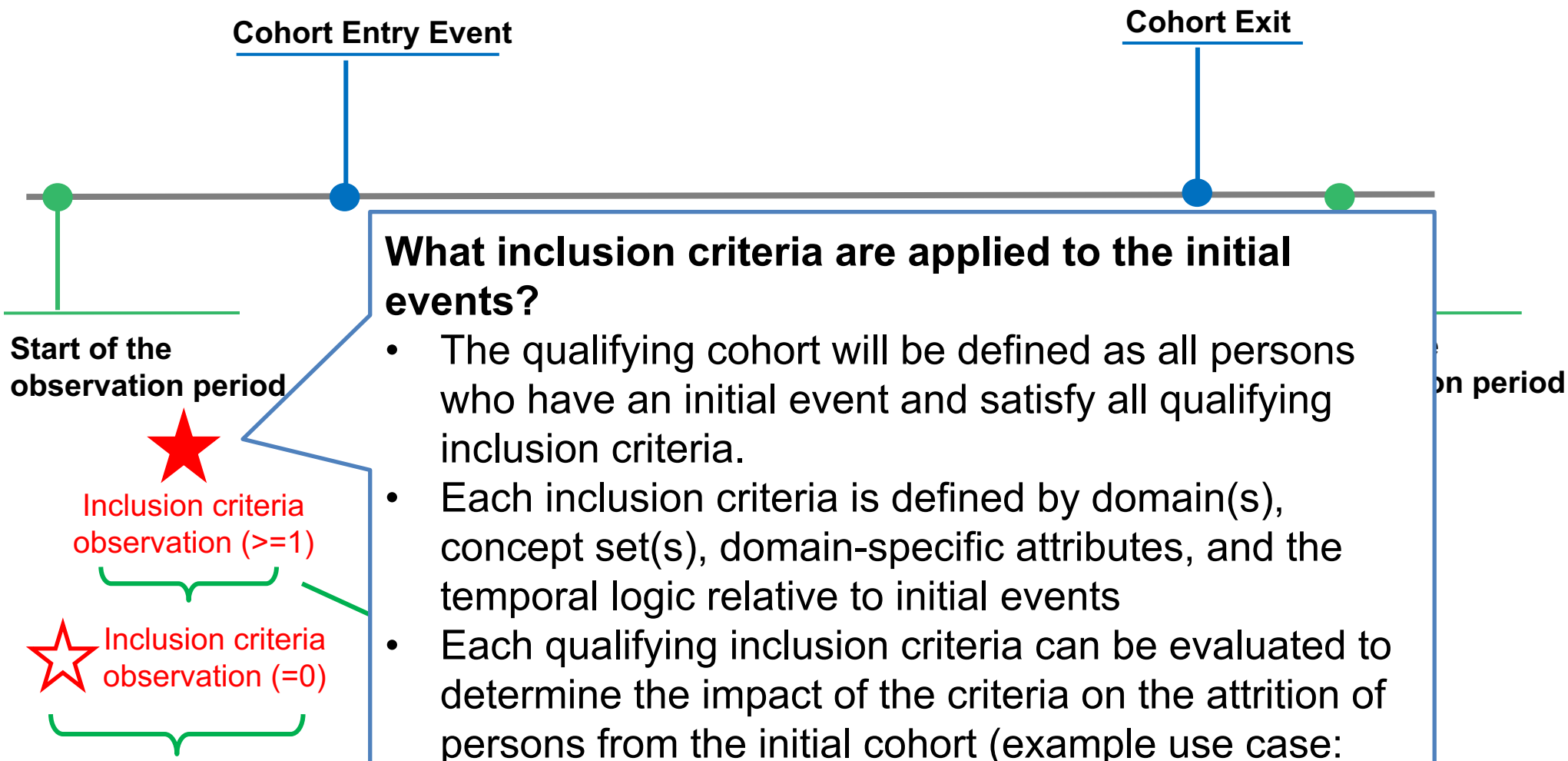
The Anatomy of a Cohort Definition



What initial event(s) define cohort entry?

- **Do:**
  - Define by **existence** of any observation in any domain
- **Don't:**
  - Define by absence of an observation - when does absence occur?
  - Define by age- year of birth is constant, but requires index date to anchor age calculation
- **Caution:**
  - Defining a cohort by calendar date can cause observation bias, since that date unlikely to be at point of health service utilization, ex: cases matched to controls. Consider instead defining by a visit that occurs within a calendar timeframe.

The Anatomy of a Cohort Definition



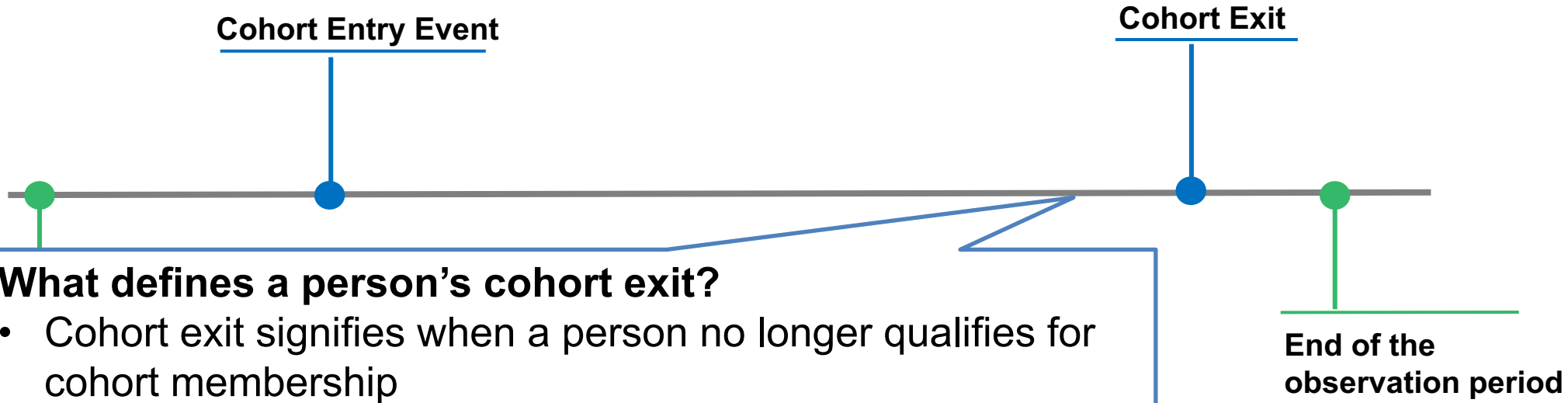
What inclusion criteria are applied to the initial events?

- The qualifying cohort will be defined as all persons who have an initial event and satisfy all qualifying inclusion criteria.
- Each inclusion criteria is defined by domain(s), concept set(s), domain-specific attributes, and the temporal logic relative to initial events
- Each qualifying inclusion criteria can be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort (example use case: clinical trial feasibility)

What inclusion criteria are applied to the initial events?

- **Do:**
  - Specify all criteria as inclusion criteria to avoid confusion of Boolean logic around inclusion vs. exclusion
  - Use information on or before index event
    (think like a randomized trial: index event is study start, can't predict future)
- **Don't:**
  - Assume temporal logic, but always provide relative time window to evaluate criteria
- **Caution:**
  - There's a difference between 'first time in history with >365d prior observation' vs. 'no prior observation in last 365 days'
  - One person may have multiple initial events, criteria are applied to each event (not person)

The Anatomy of a Cohort Definition



What defines a person's cohort exit?

- Cohort exit signifies when a person no longer qualifies for cohort membership
- Cohort exit can be defined in multiple ways:
  - End of observation period
  - Fixed time interval relative to initial event
  - Last event in a sequence of related observations (ex: persistent drug exposure)
  - Censoring observations
- Cohort exit strategy will impact whether a person can belong to the cohort multiple times during different time intervals

What defines a person's cohort exit?

- **Do:**
  - Specify a cohort exit, even if you are not intending to use it for your analytic use case
- **Don't:**
  - Confuse censoring for analytical purposes with cohort definition (which can be analysis-independent)…ex: censoring at time of outcome
- **Caution:**
  - Time-of-cohort participation can be different from analysis time-at-risk…ex: acute effects can be studied using a fixed window post-exposure start, intent-to-treat analysis can follow person through observation period end

Defining Cohorts in Atlas

1 Questions to answer when defining a cohort

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?

2 Cohort components

- Cohorts are defined using
  - Domain(s)
  - Concept set(s)
  - Domain-specific attributes
  - The temporal logic relative to initial events

Concept sets

Concept Set Expressions

- **Concept Set:** logical expression to represent a list of concepts in the OHDSI vocabularies encompassing a clinical entity of interest
  - List of one or more concepts
  - Optional operator for each concepts in the list:
    - **Exclude**: Exclude this concept (and any of its descendants if selected) from the concept set.
    - **Descendants**: Consider not only this concept, but also all of its descendants.
    - **Mapped**: Allow to search for non-standard concepts.

- **Concept Set** can be thought of as a standardized, computer-executable equivalent of the code lists often used in observational studies.
- A concept set expression can be materialized into a list of concepts using any instance of the OHDSI vocabularies
  - JSON expression executed via webAPI into standard SQL query

What to look for?

Face validity: Atlas

Based on your data

1. Included vocabularies: do we see all the vocabularies we expect?
2. Standard concepts: do we see the type of concepts we expect (standard/source/classificational)?
3. Do we see the concepts we want among the most common concepts?

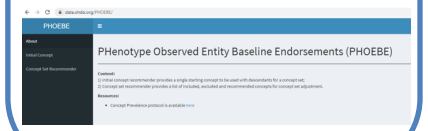
Recommender system: PHOEBE

Based on the records collected across the network

1. Check included concepts
2. Check non selected descendants
3. Check non selected parents
4. Check recommended concepts (recommended through standard and recommended through source)

Phenotype development and evaluation workflow

Concept set expressions



Cohort definition logic

Initial events

Conceptsets

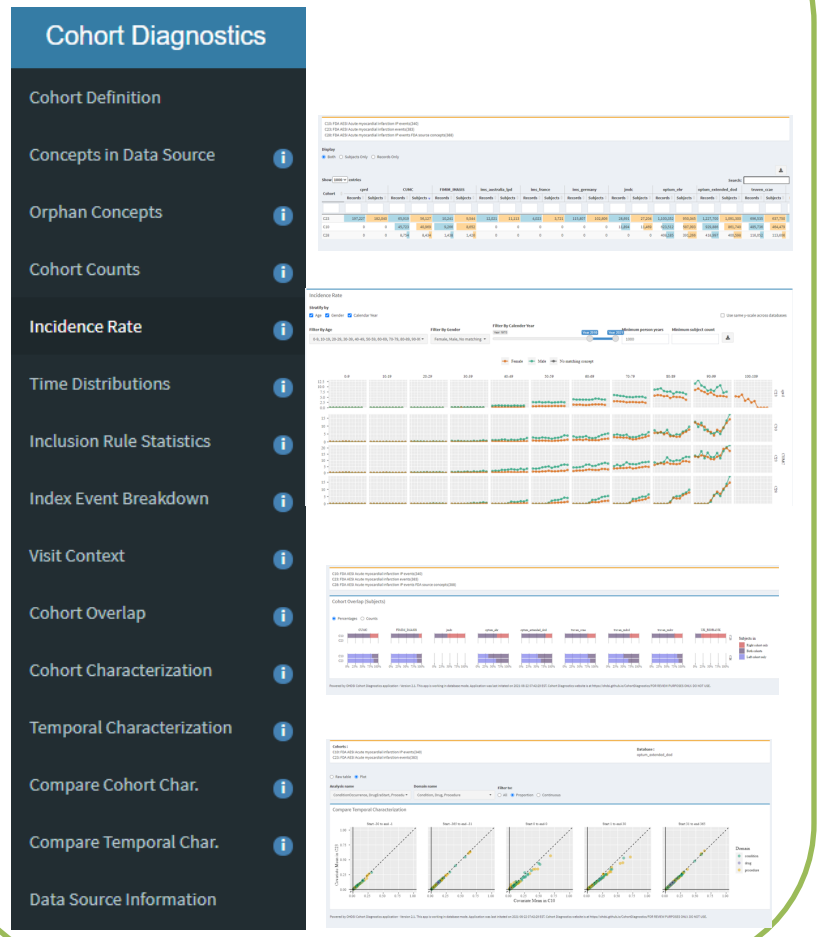
Inclusion criteria

Conceptsets

Exit strategy

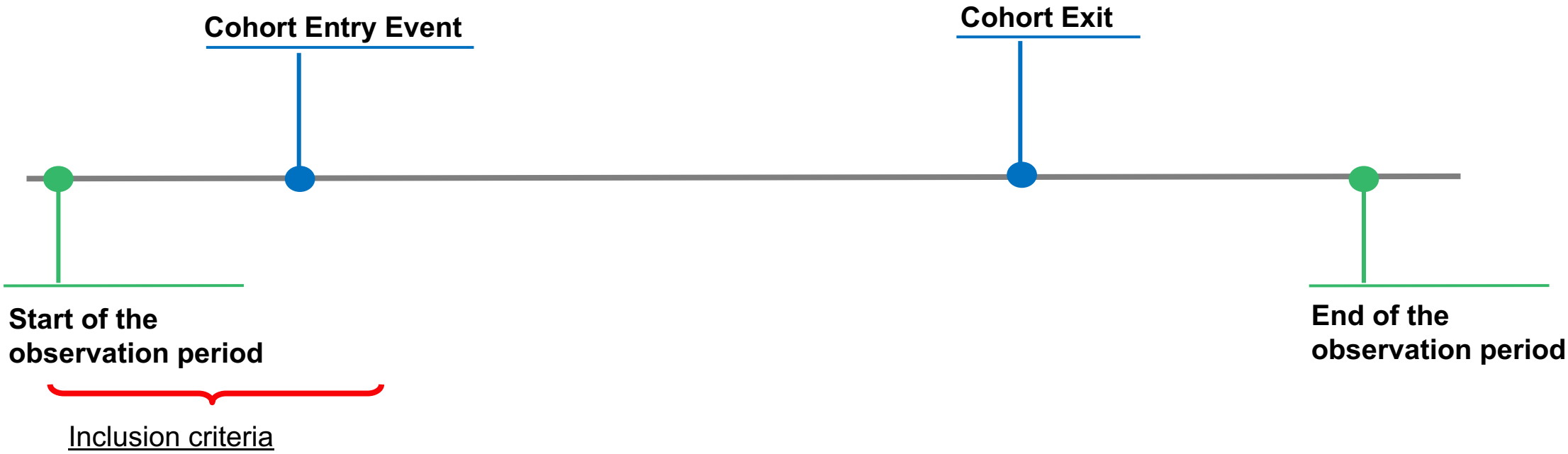
Conceptsets

Cohort diagnostics



Create cohort for ‘new users of lisinopril with prior hypertension’ in ATLAS

The Anatomy of a Cohort Definition



Defining the “new users of lisinopril with prior hypertension” Cohort

Inclusion criteria:

- History of hypertension **one year prior**