Experimental Design with the HCUP SID

Using Administrative Data for Clinical and Health Services Research

Goal

- Explore generic study designs for SID data

- Go from full SID data to subset containing *primary inclusion criterion*

    - Usually consists of diagnoses/procedures

    - From this smaller data set, more specific inclusion/exclusion criteria can be applied, and that information can be reported in a cohort flow diagram
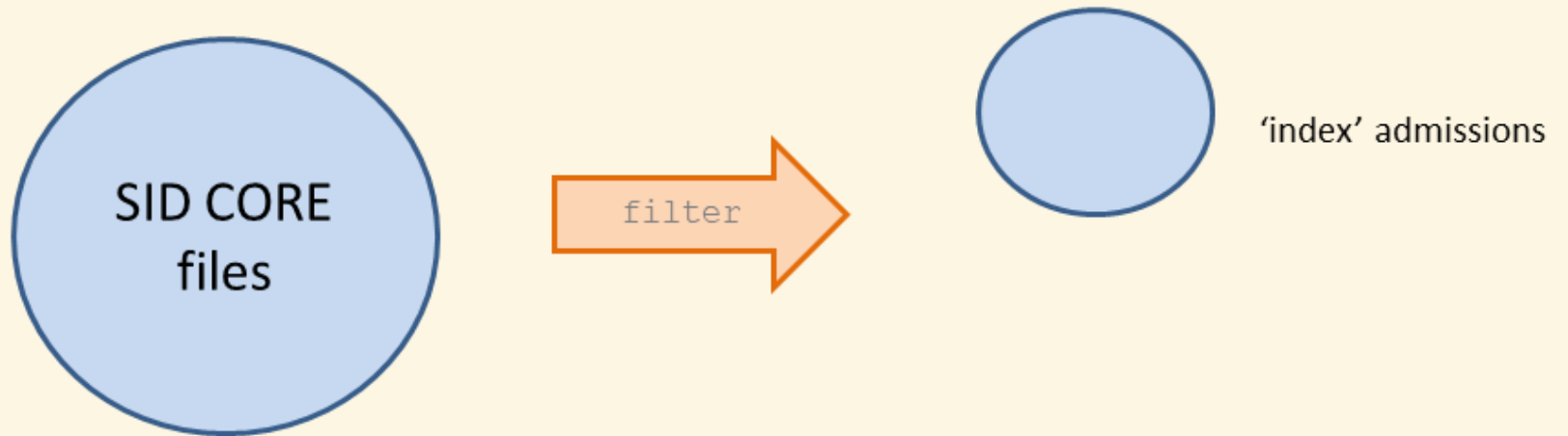
Patient- vs Admission-level

- What is the unit of analysis?

- Admission-level analysis uses all records that meet inclusion/exclusion criteria

  - Cannot define baseline covariates / outcomes outside of index admissions

  - Similar to an analysis using the NIS database

- Patient-level analysis starts with all records that meet inclusion/exclusion criteria but then chooses one record per patient, typically the earliest

  - Can use other records associated with these patients for baseline covariates / outcomes
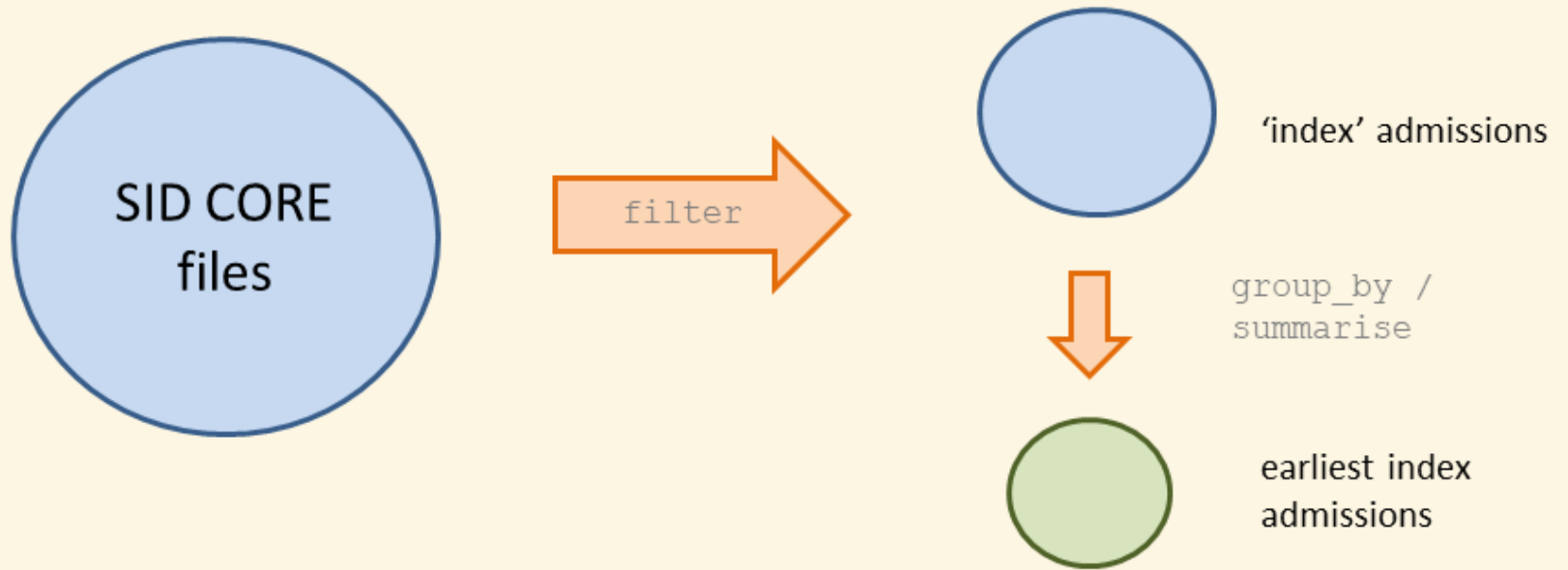
SID CORE
files
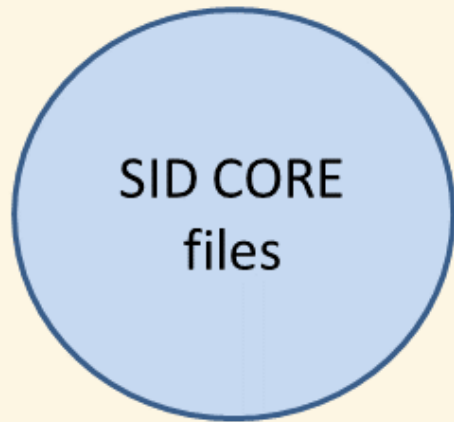
◻ discharge-level

◻ patient-level

◻ data processing step

Admission-level Design

SID CORE files

filter →

'index' admissions

Legend:
- discharge-level
- patient-level
- data processing step

Patient-level Design

SID CORE files

filter →

'index' admissions

group_by / summarise ↓

earliest index admissions

- ☐ discharge-level
- ☐ patient-level
- ☐ data processing step

Patient-level Design
(with baseline, index, and/or follow-up periods)

SID CORE files

filter

'index' admissions

group_by / summarise

earliest index admissions

inner_join

'all records' for patients in index sample
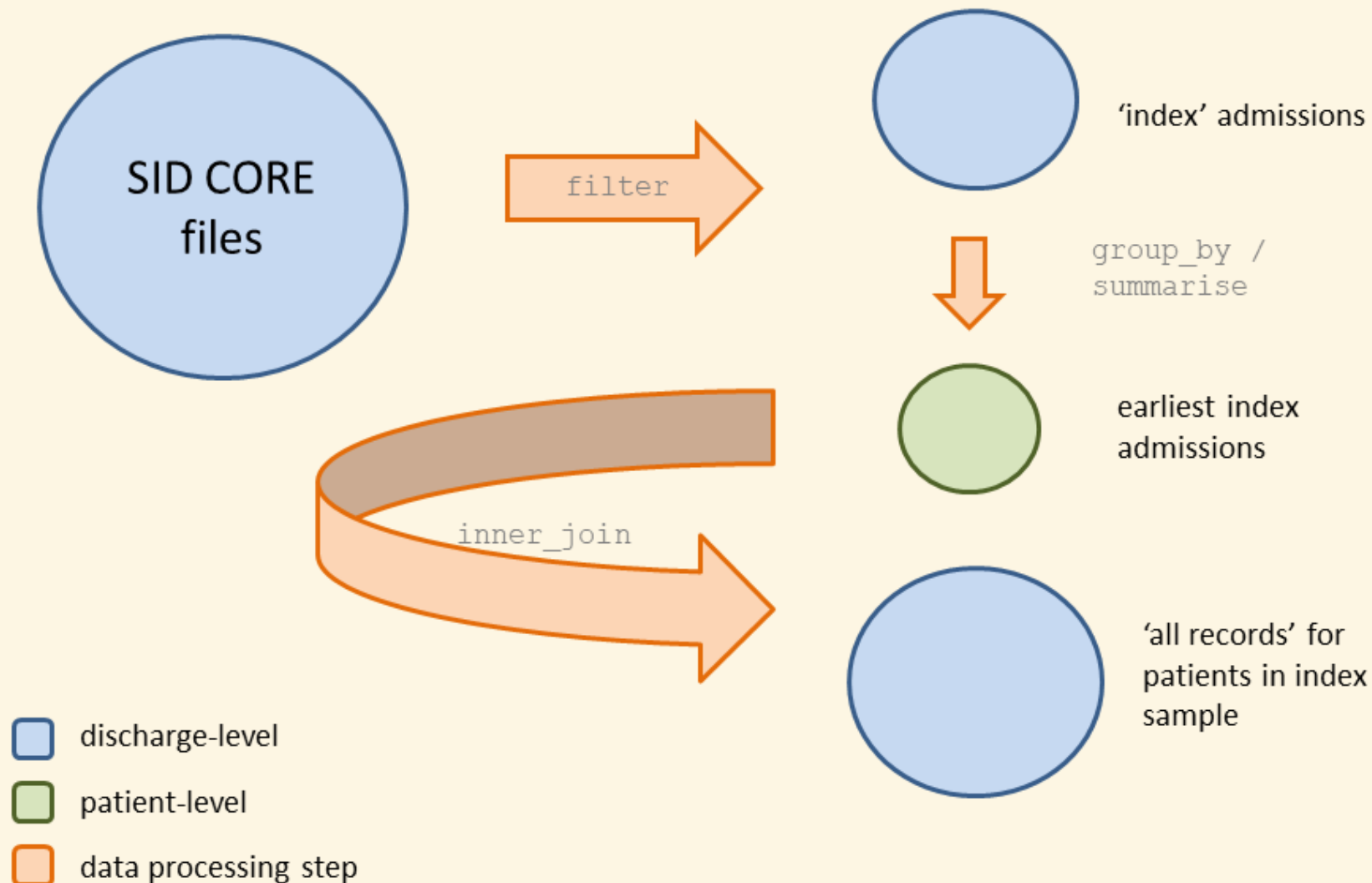
- discharge-level
- patient-level
- data processing step

Patient-level Design
(with baseline, index, and/or follow-up periods)

'all records' for patients in index sample

filter

baseline records

index records

follow-up records

- ▢ discharge-level
- ▢ patient-level
- ▢ data processing step
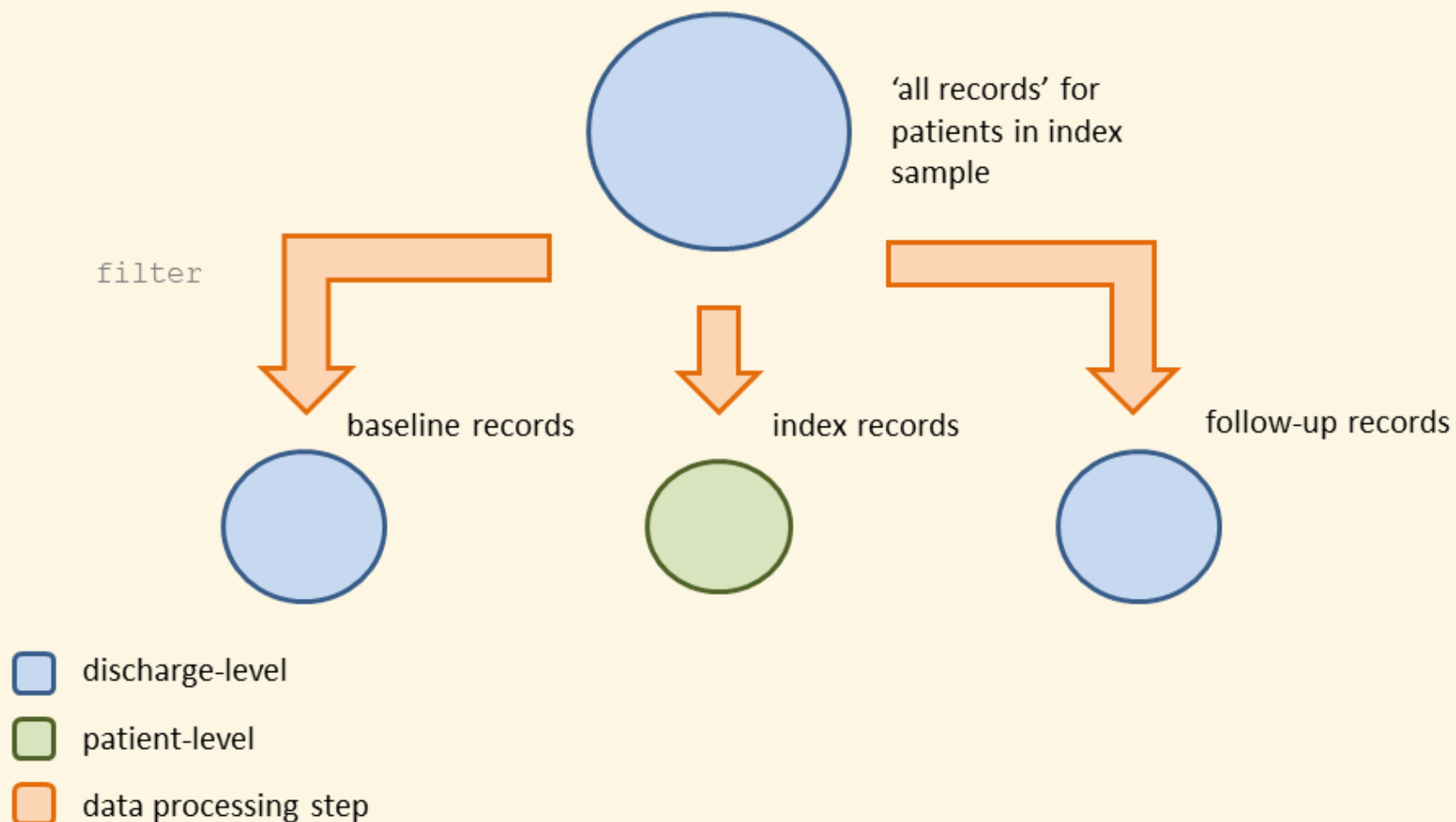
Considerations

- Choice of design is influenced by multiple factors

- Admission-level design is more straightforward to code

    - Individual records do not provide much information about timing of events

    - Limits the types of conclusions that can be made

- (Longitudinal) patient-level design requires more coding

    - More planning required to keep track of multiple data sets

    - Allows a larger set of potential outcomes