

**OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP**

**Standardized Tools for Data
Characterization and Utility of
Exploratory Visualization**

Paul Stang, Bram Hartzema, Patrick Ryan
on behalf of OMOP Research Team
January 11, 2011

Overview

- A detailed, high-level overview of the characteristics of the databases used in OMOP
- Standardized methods, e.g. OSCAR and NATHAN, have been developed in OMOP to characterize data sources
- Visualizations of data using Spotfire can provide additional insights and is useful in the interpretation of findings
- **Premise:** The OMOP drugs of interest and health outcomes of interest are made up of a large set of drugs and conditions, each of which could have unique behavior that should be considered. We need a comprehensive view of these populations.

Main Problem to 'Solve'

- No longer working within a single database/single study so the basic 'demographics' of a *Table 1* will not suffice
 - Neither will the spot checks for anomalies
- We need to know the underlying data characteristics across the surveillance network in order to understand any differences that may arise across a surveillance system
- This can be complicated and tedious and is inclusive of some data 'quality' issues

OSCAR & NATHAN

OSCAR (Observational Source Characteristics Analysis Report)

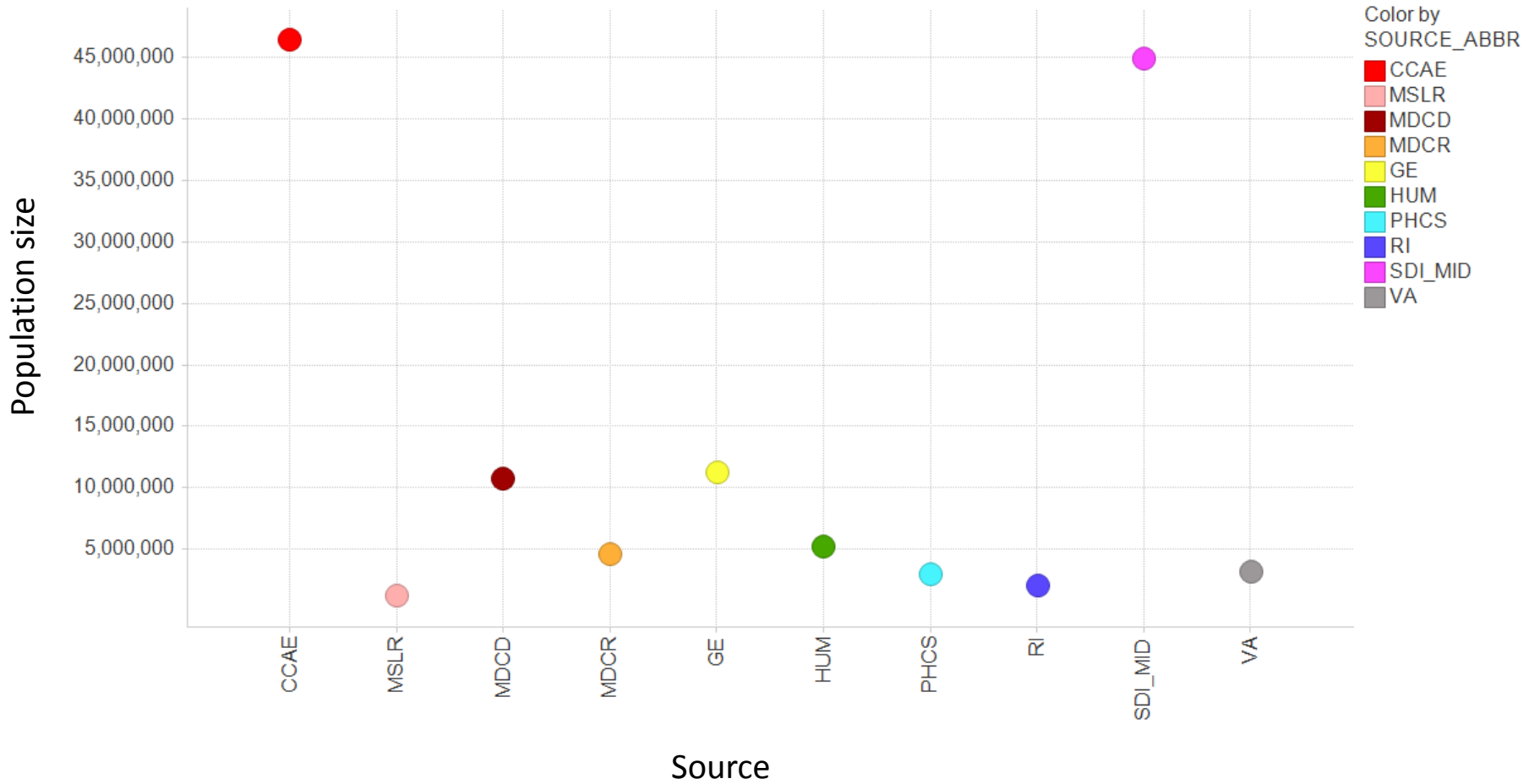
- An SAS program that creates a comprehensive set of over 1 million descriptive statistics
 - summarizes available data from a given source within the OMOP CDM;
 - provides context for interpreting and analyzing findings of drug safety studies;
 - facilitates comparisons between data sources;
 - inclusive of spot checks proposed by Hennessy, HMORN;
 - enables comparison of overall database to specific subpopulations of interest;
 - supports validation of transformation from raw data to OMOP CDM.

An extension of OSCAR is NATHAN (Natural History Analysis), which provides summaries for subpopulations of interest

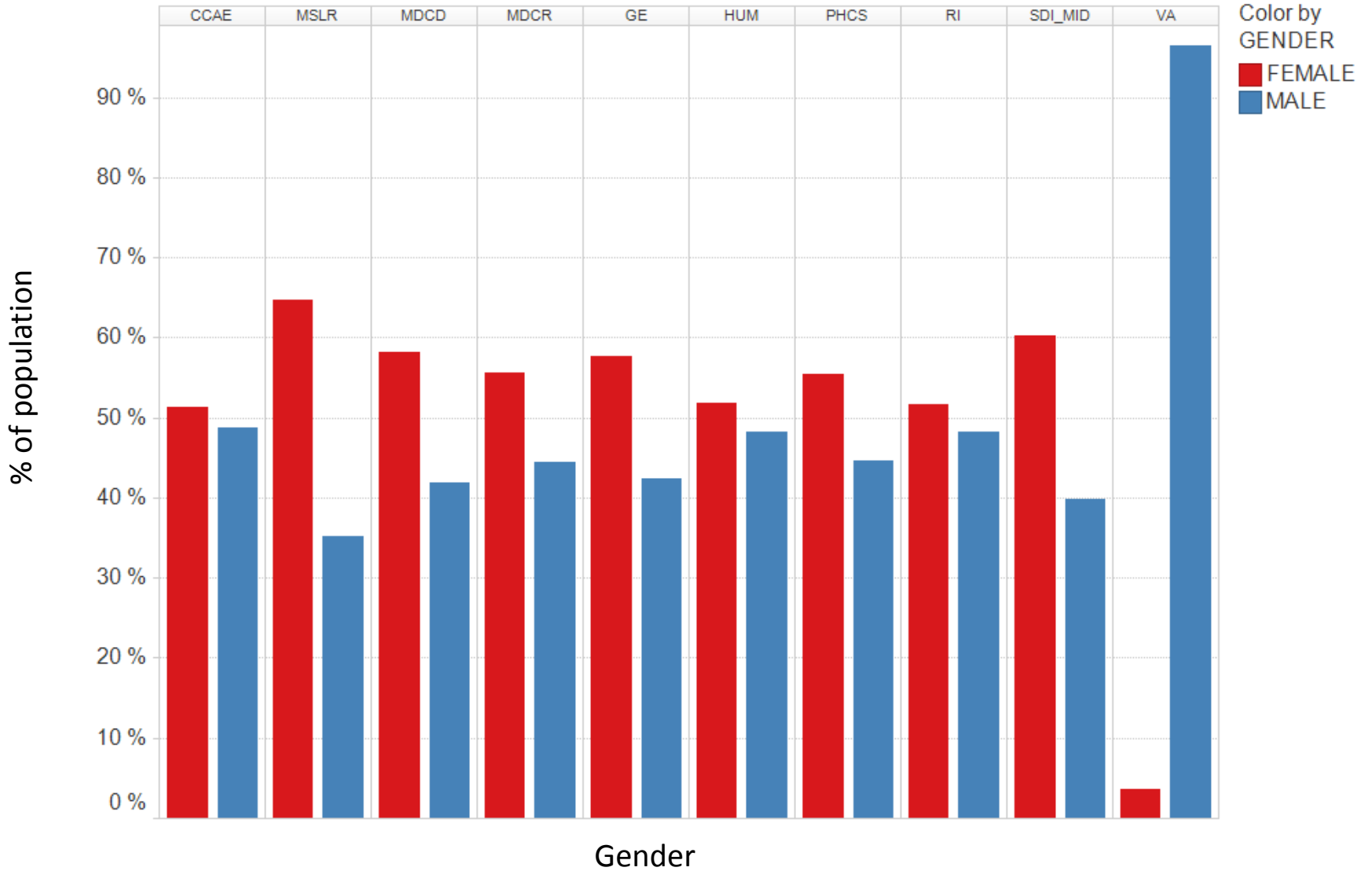
OMOP Data Characteristics

DEMOGRAPHICS

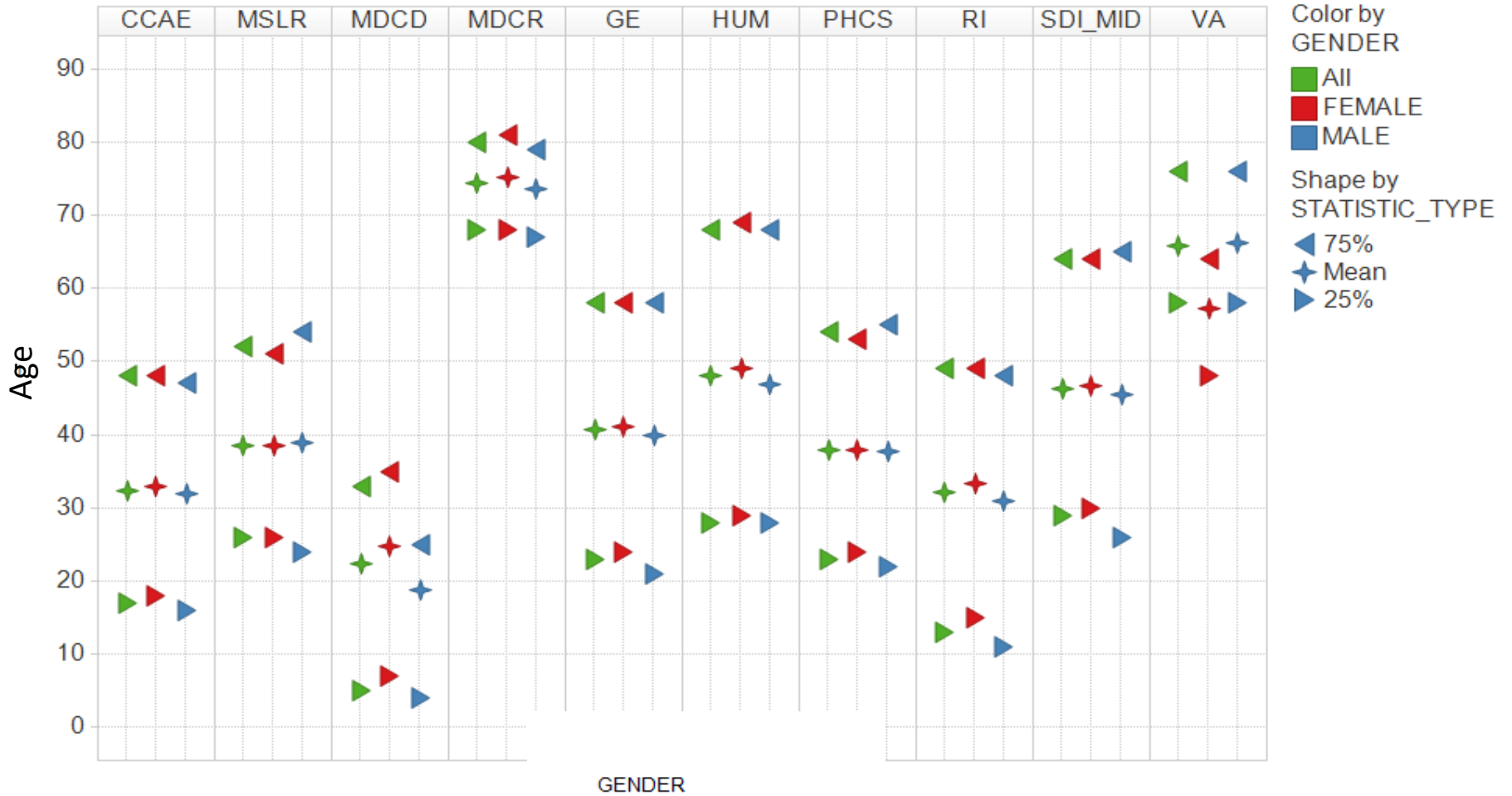
OMOP data community - population size



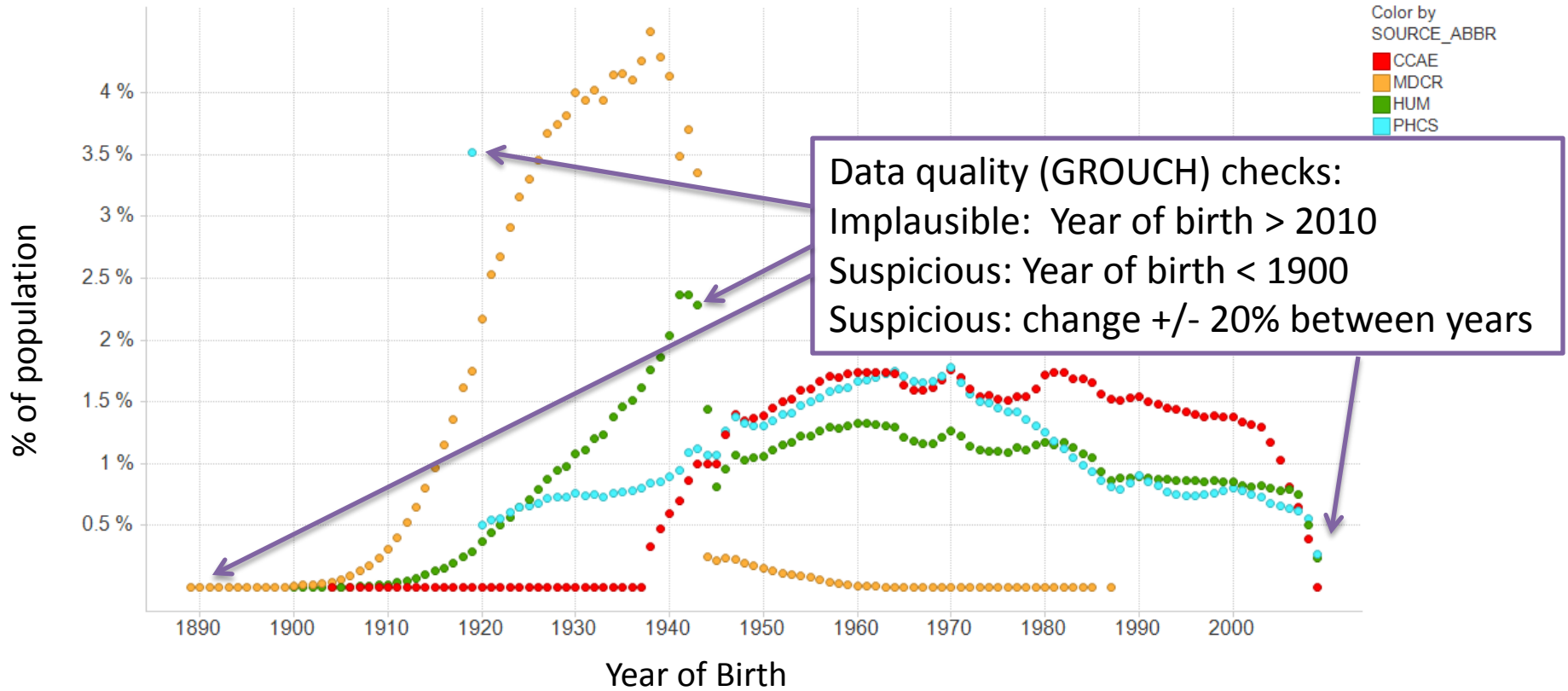
Sources by Gender



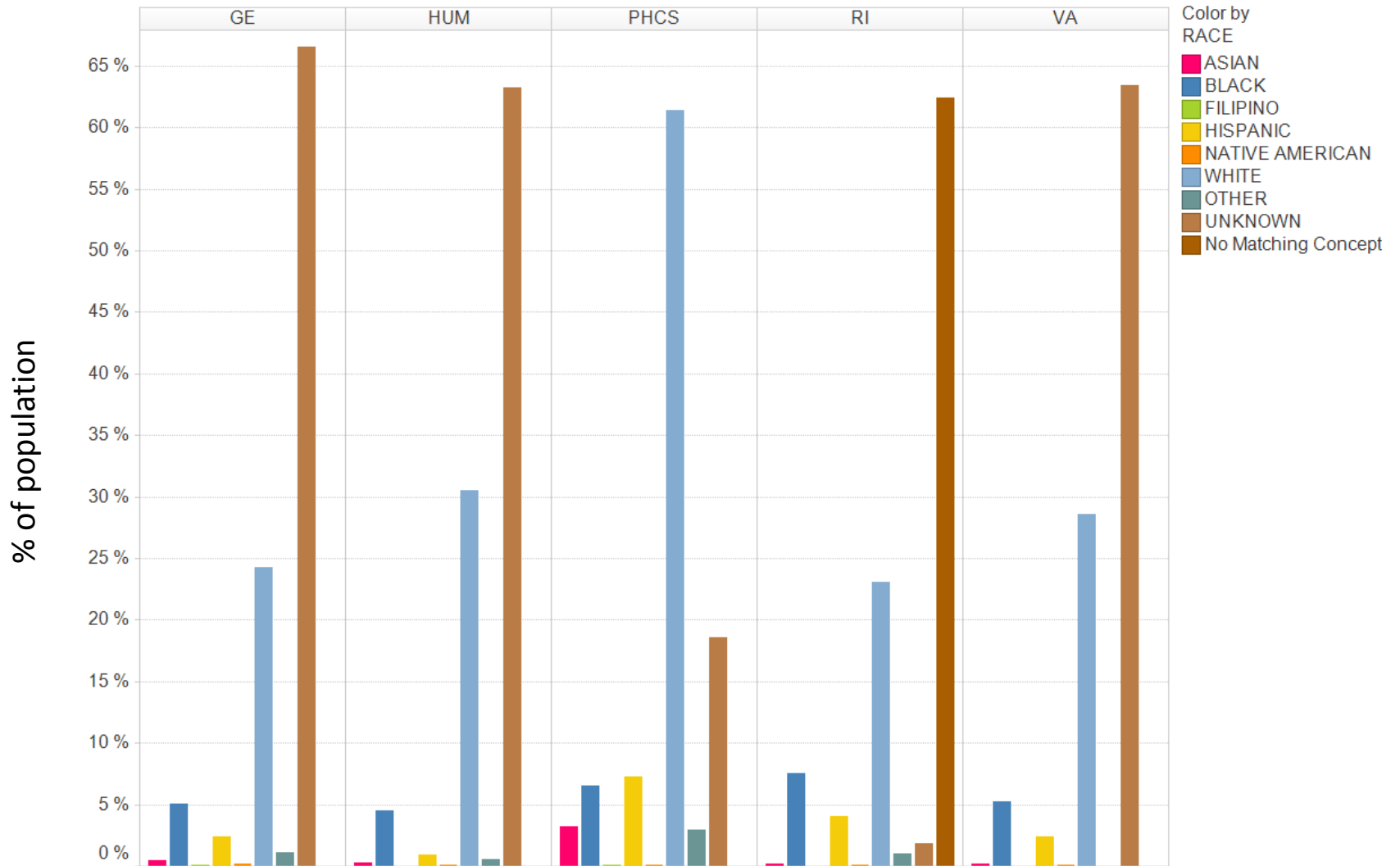
Sources by age distribution



Age distribution



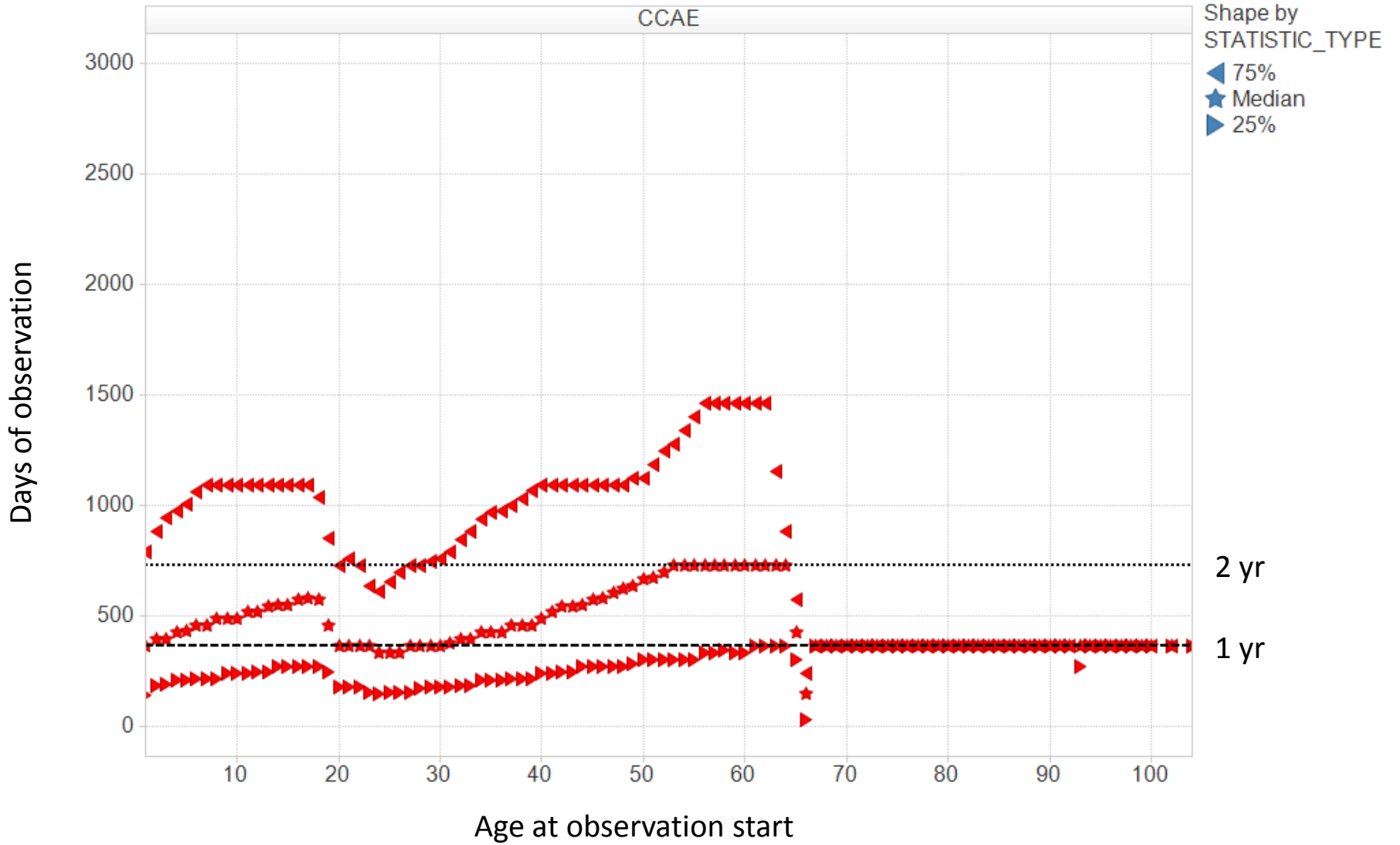
Race distribution



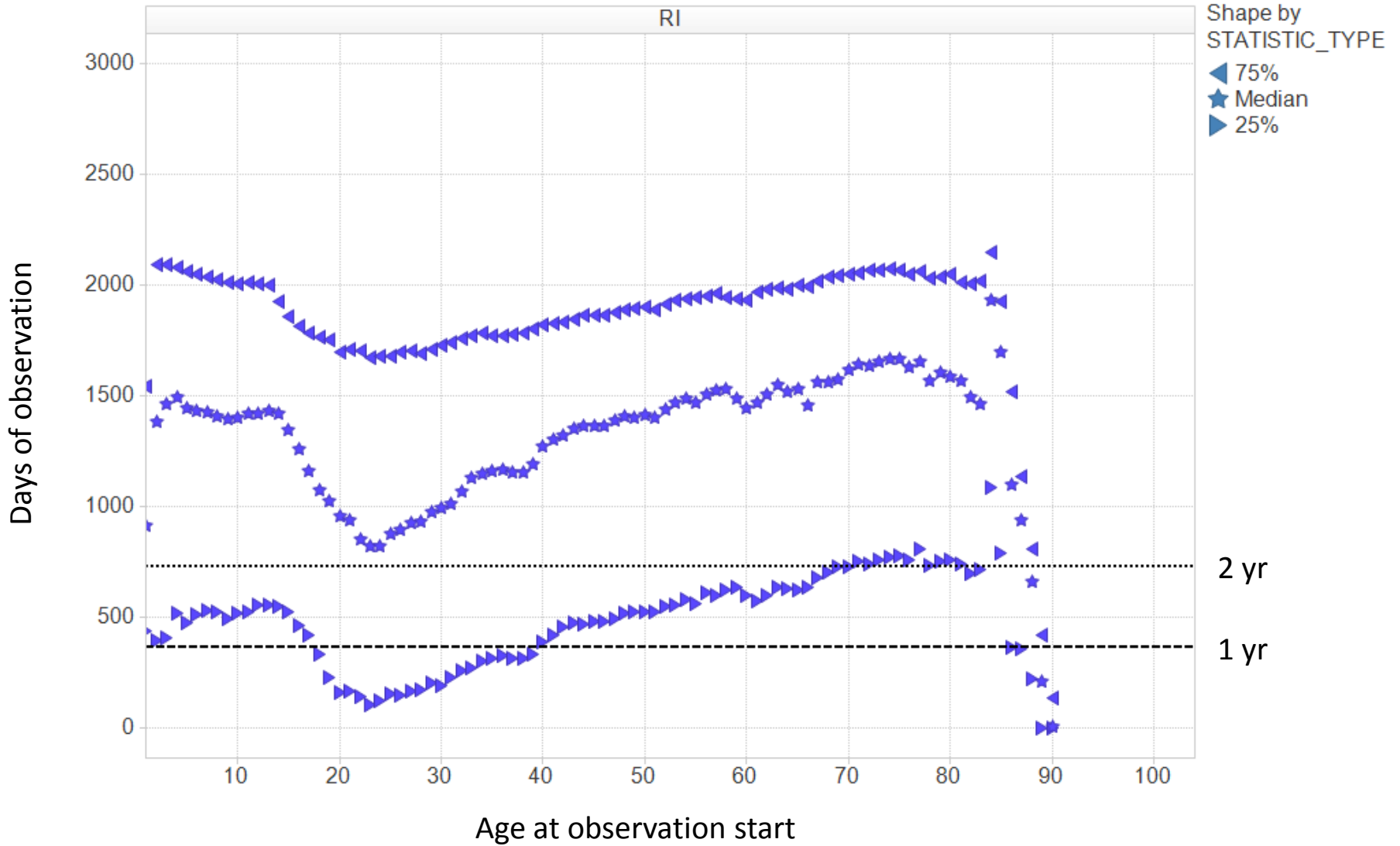
OMOP Data Characteristics

LONGITUDINAL FEATURES

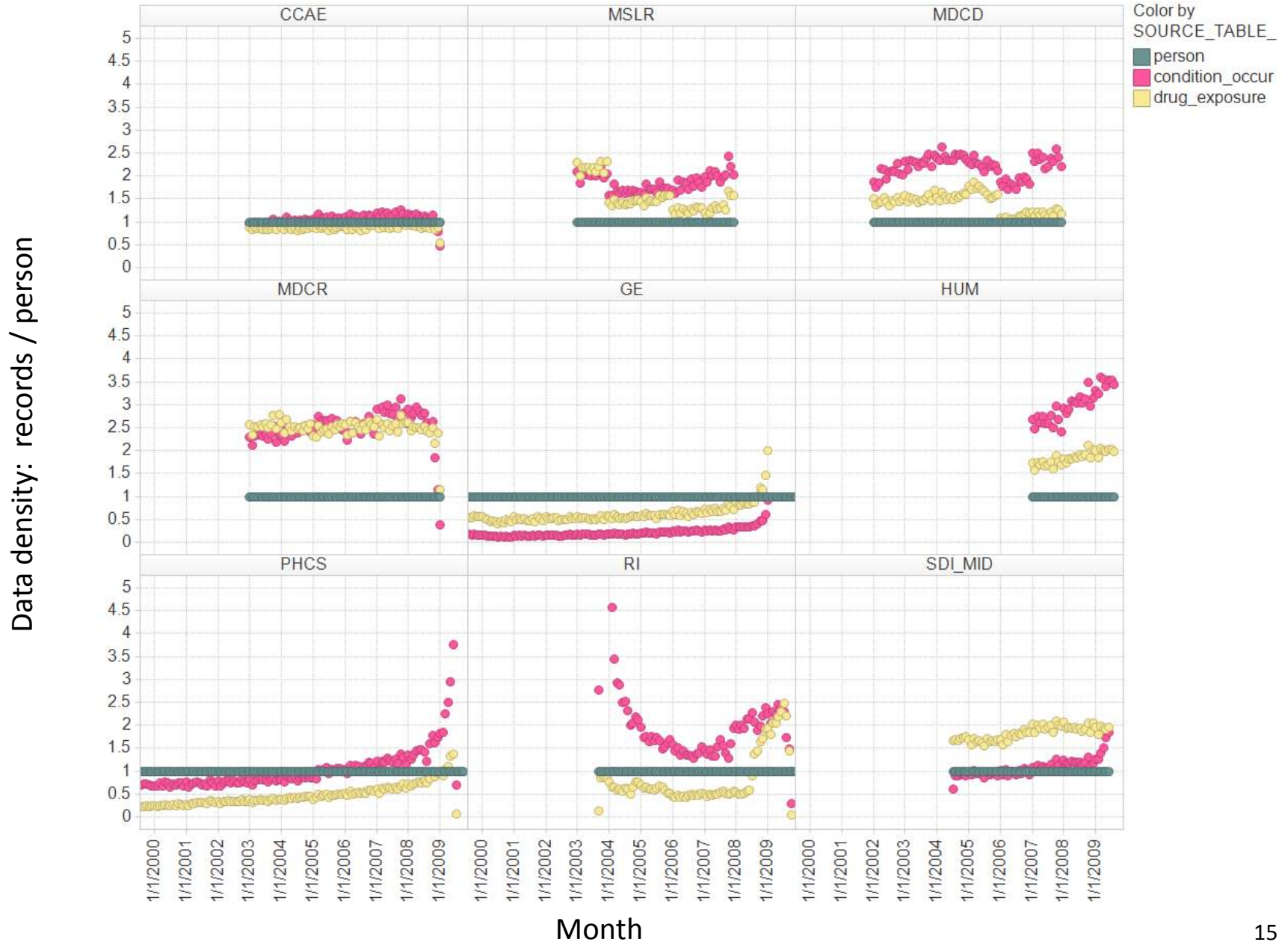
Observation period length



Observation period length



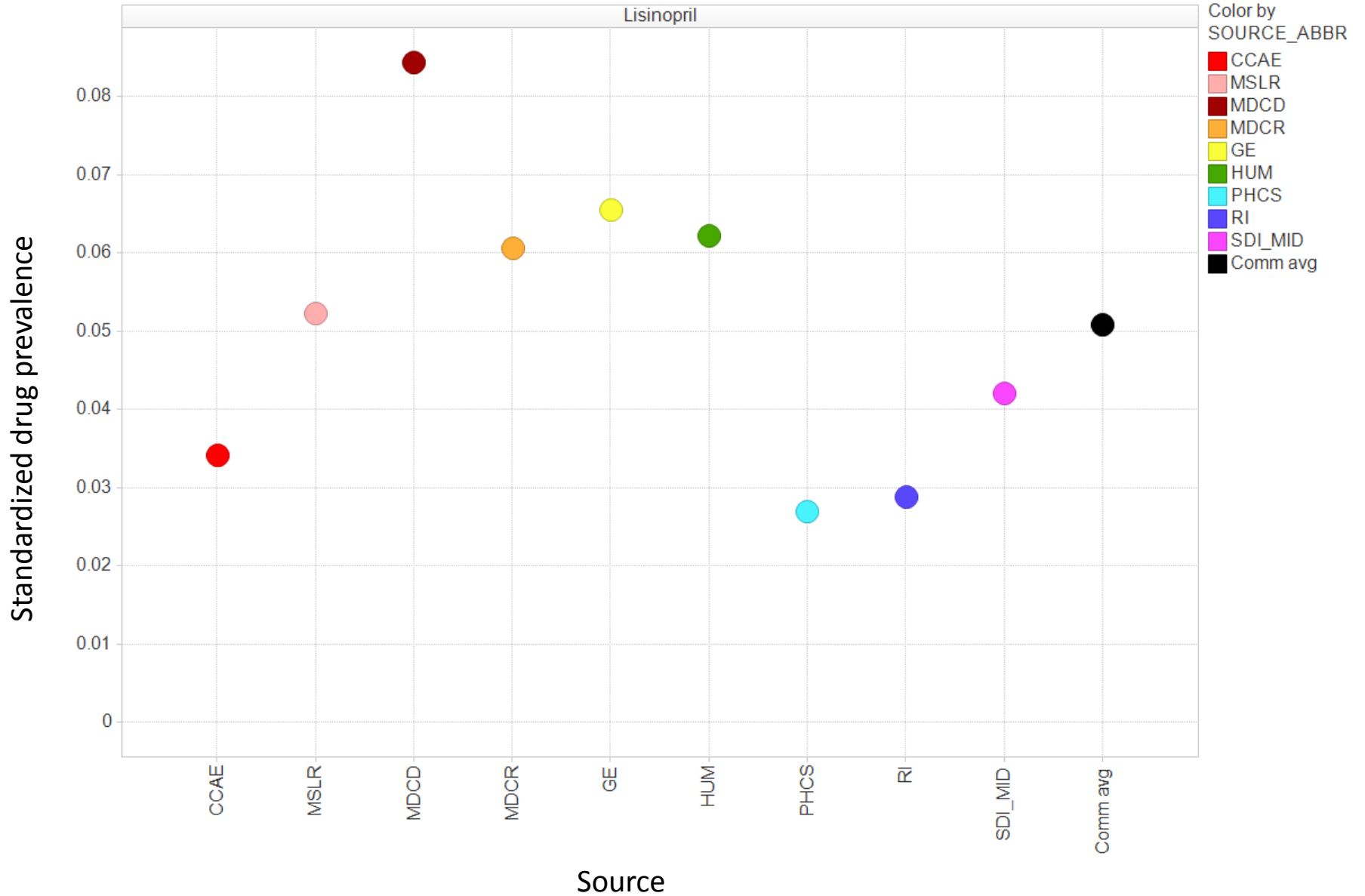
Records per person over time



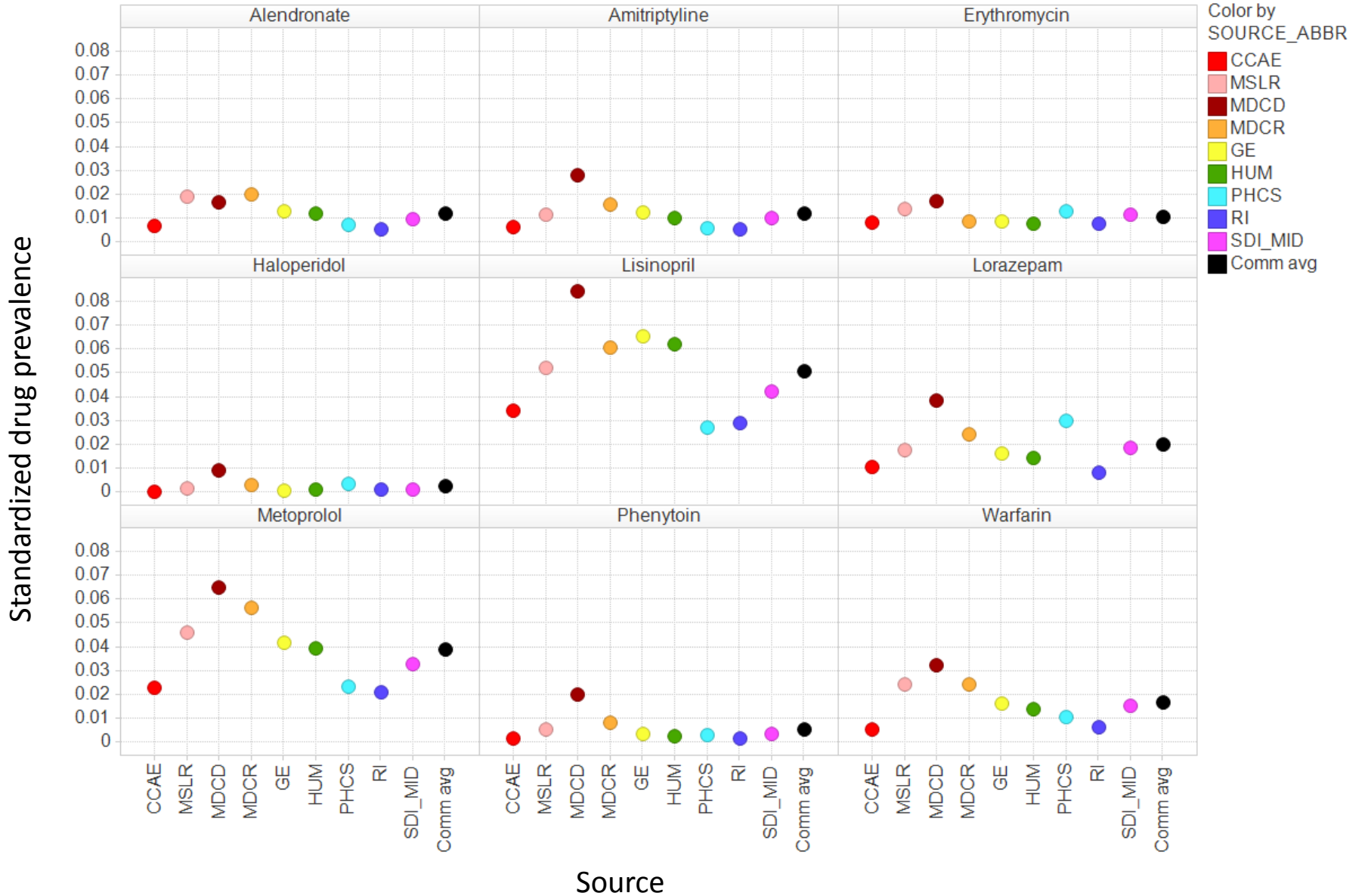
OMOP Data Characteristics

PREVALENCE AND CHARACTERISTICS OF MEDICATION USE

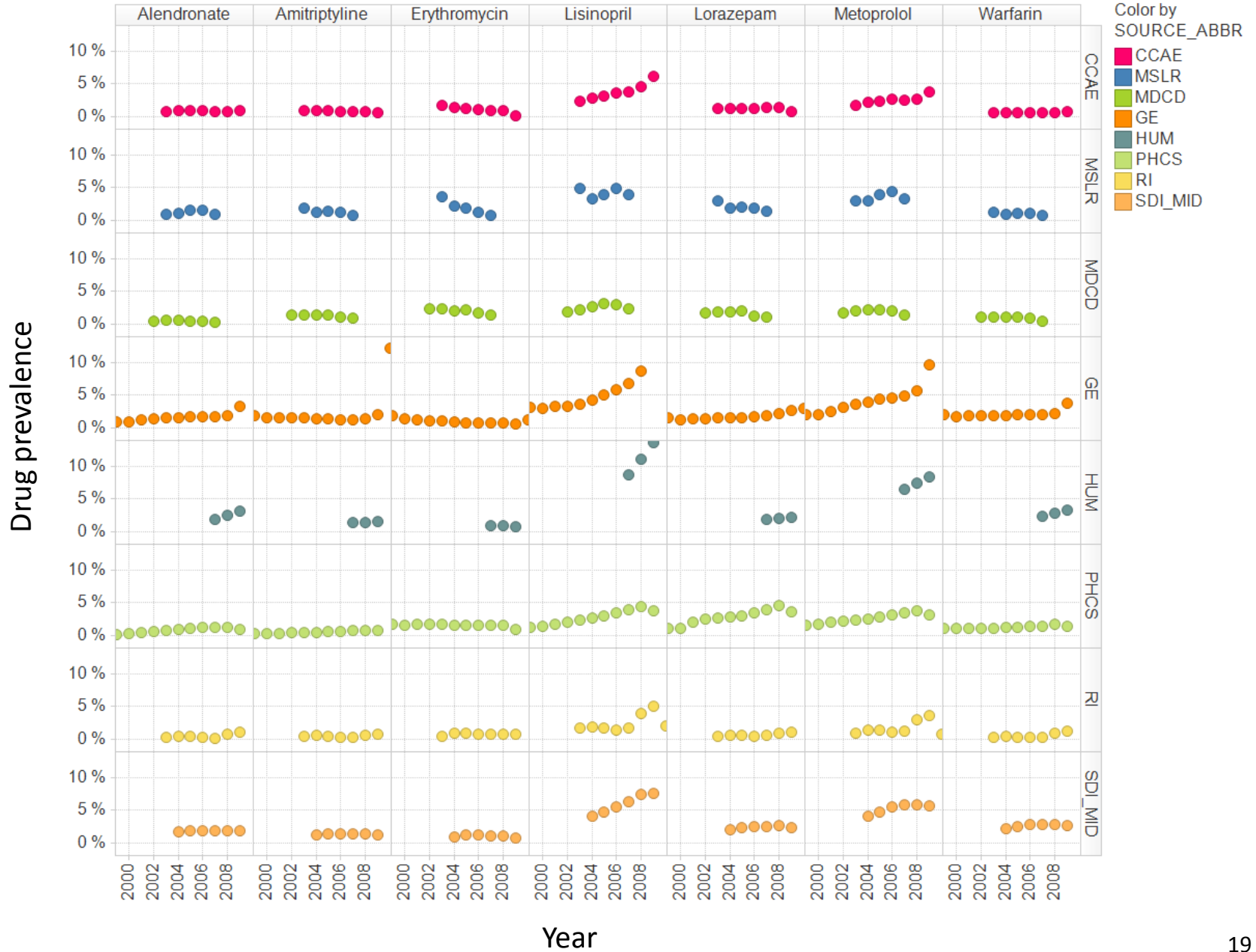
Standardized drug prevalence



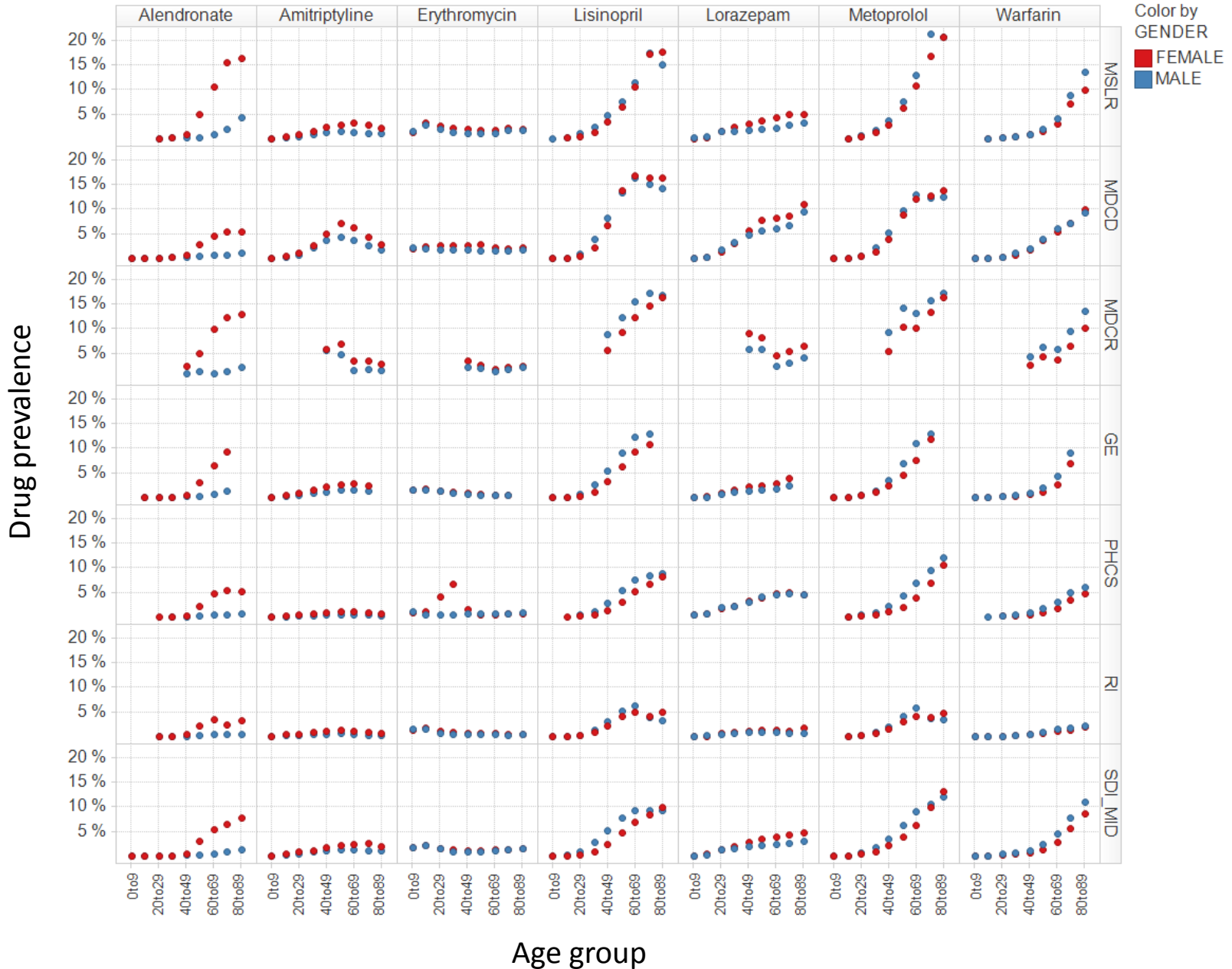
Standardized drug prevalence



Drug prevalence by year



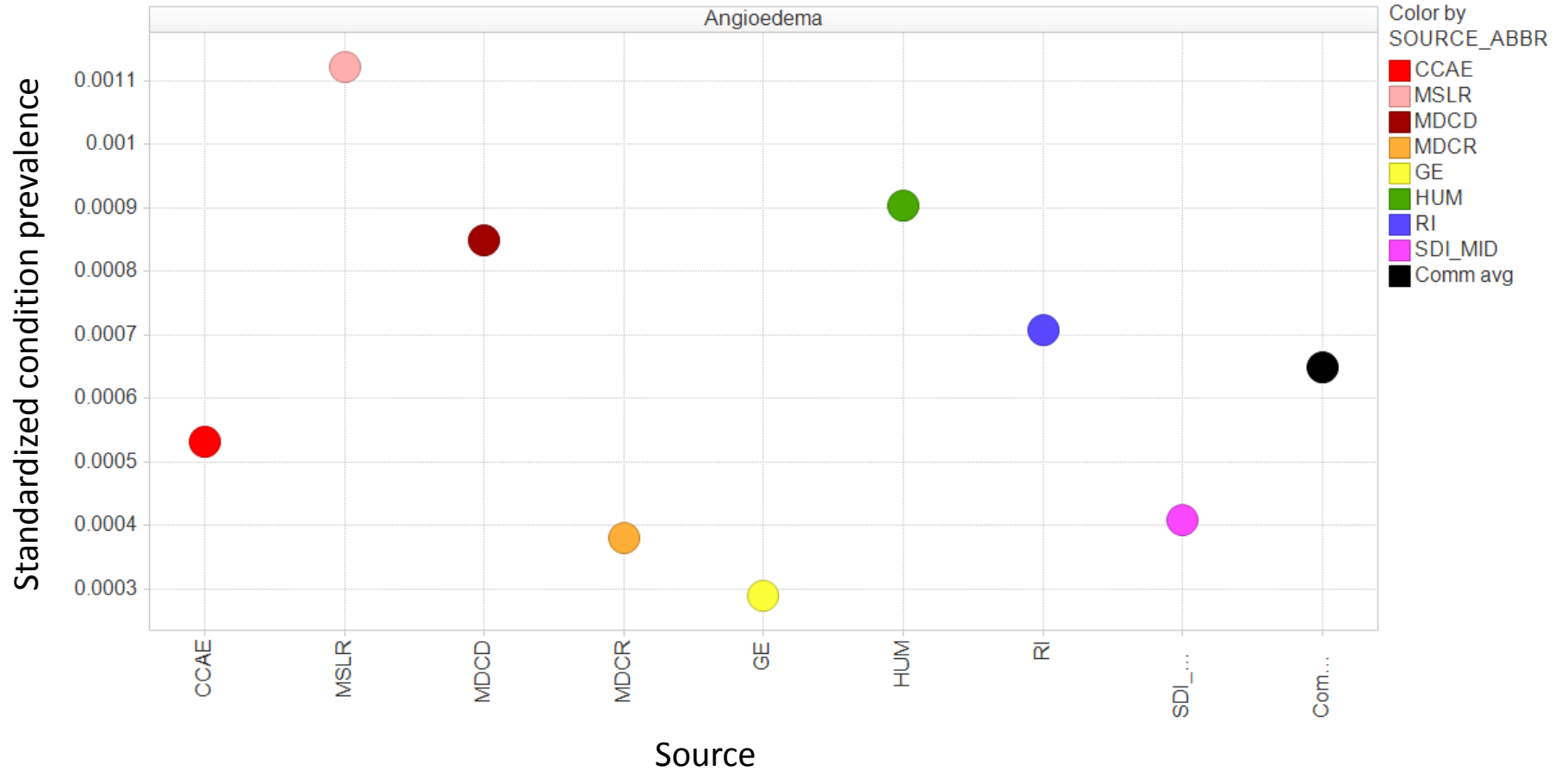
Stratified drug prevalence by age group



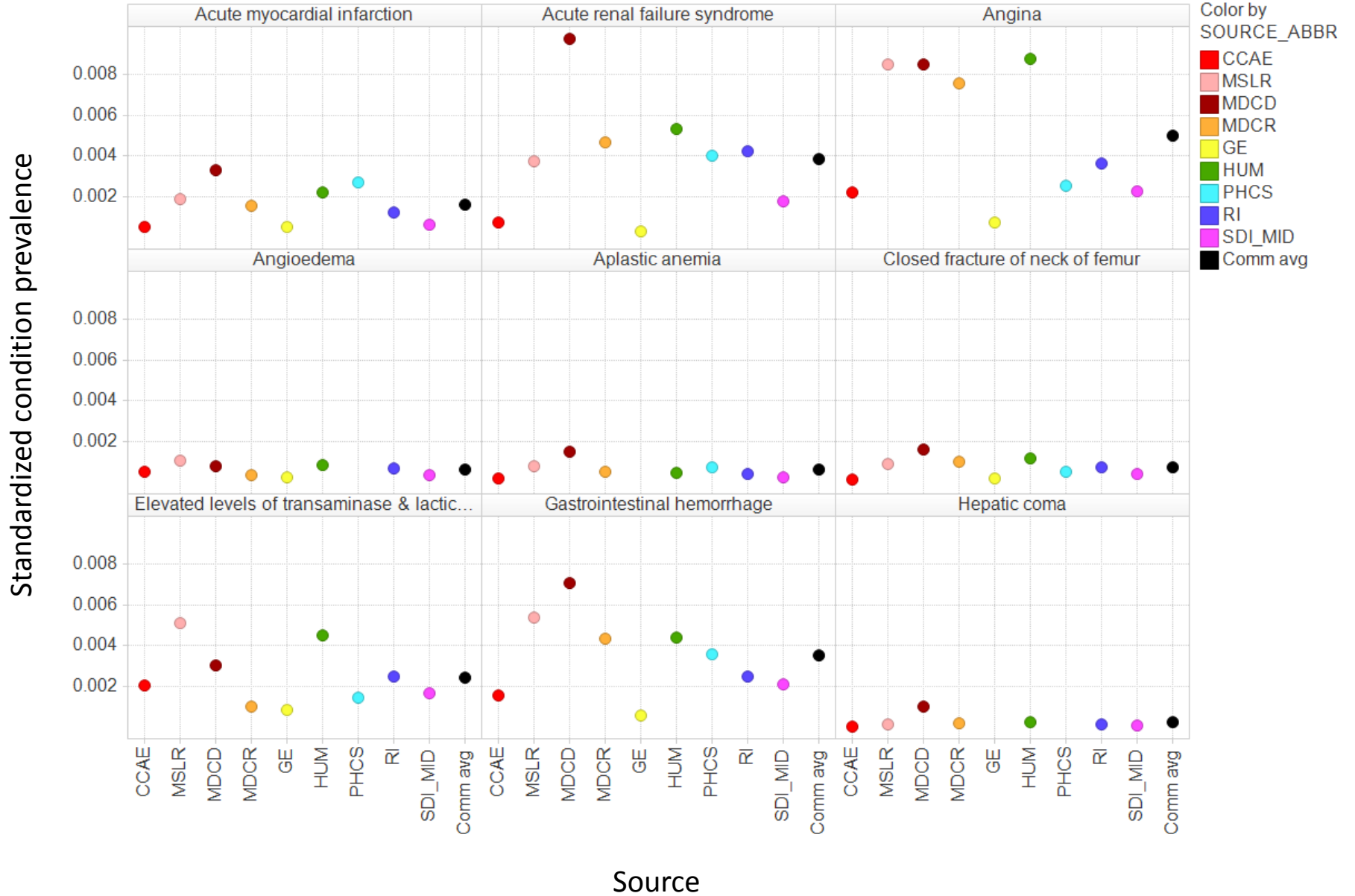
OMOP Data Characteristics

CHARACTERISTICS OF CONDITIONS

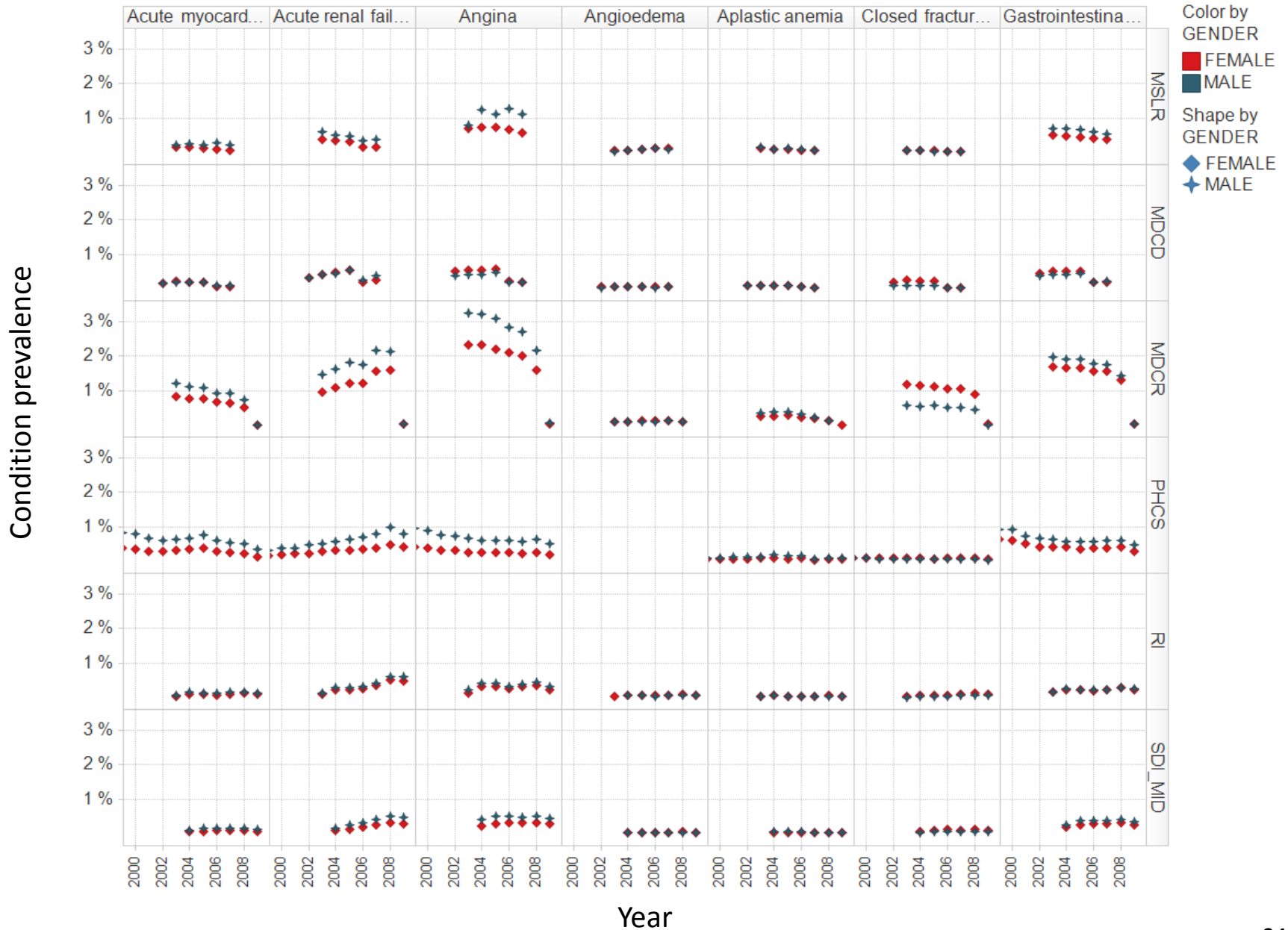
Standardized condition prevalence



Standardized condition prevalence



Stratified condition prevalence by year



Summary

- Detailed understanding of the characteristics of datasets is a prerequisite for active drug safety surveillance
- OMOP has developed the tools to facilitate this understanding in OSCAR and NATHAN
- A network of disparate data sources provide the comparative context for understanding data characteristics and increases the generalizability of the findings
- Visualizations of this massive amount of data are required to efficiently explore and interpret the findings