

Empirical Assessment of Alternative Methods for Identifying Seasonality in Observational Healthcare Data

PRESENTER: **Anthony Molinaro**

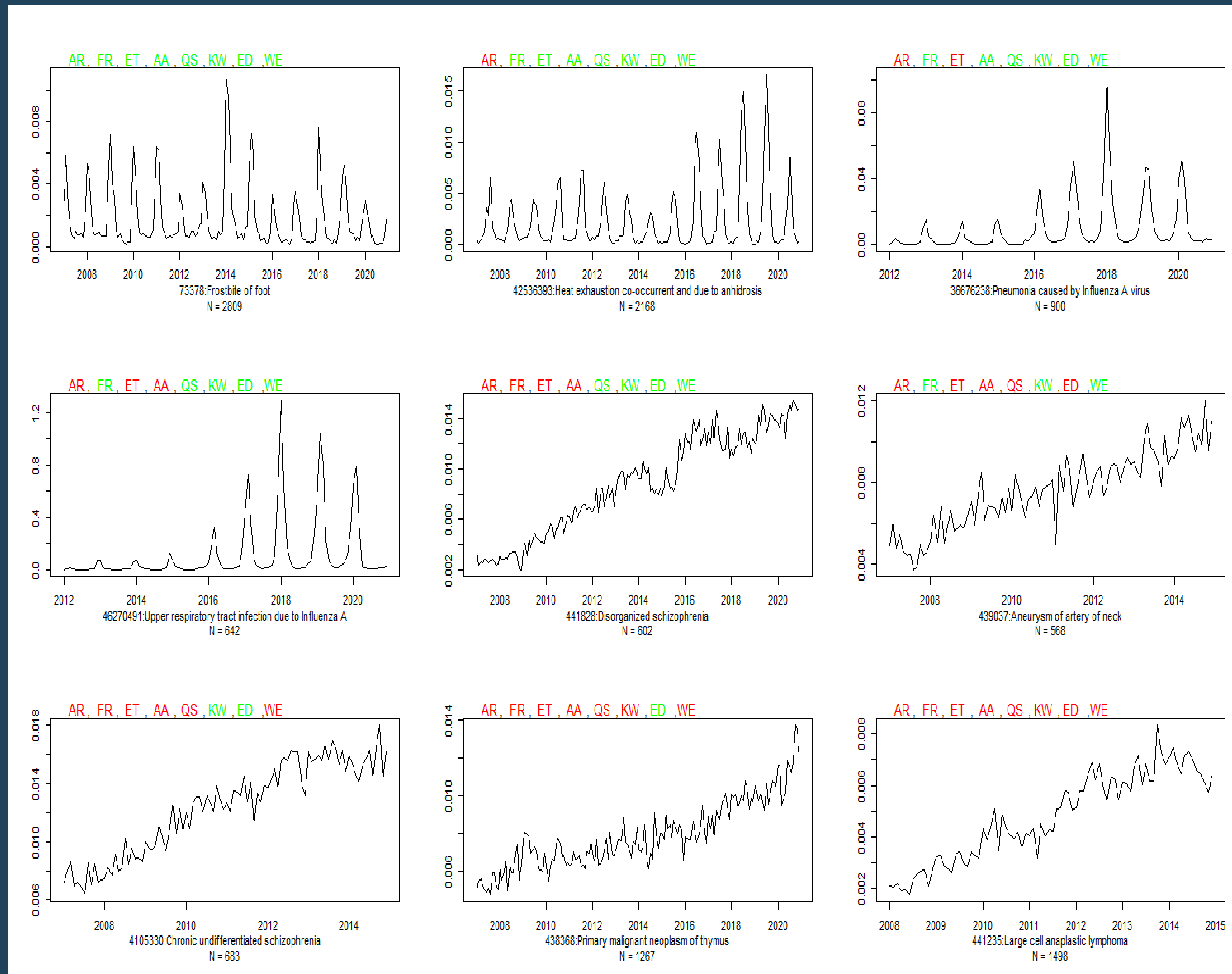
INTRO:

- Understanding the seasonality of a biological event can contribute to an improved understanding of its causes and help guide appropriate responses.
- Observational data contains timestamped diagnosis codes that can be used to create time series that serve as proxies for biological events.
- Automated methods must be employed to determine the seasonality of thousands of time series.

METHODS

- Leveraging the OHDSI packages CASTOR (<https://github.com/OHDSI/Castor>) and ACHILLES (<https://github.com/OHDSI/Achilles>), we create monthly proportion time series from condition concepts with at least four years of complete data. We compare the following eight popular methods of binary seasonality classification of time series at the .01, .05, and .1 levels of significance:
 - Friedman's Test (FR)
 - Welch's Test (WE)
 - Kruskal-Wallis Test (KW)
 - Edwards' Test (ED)
 - QS Test (QS)
 - ETS Test (ET)
 - Auto ARIMA Test (AA)
 - ARIMA Test (AR)

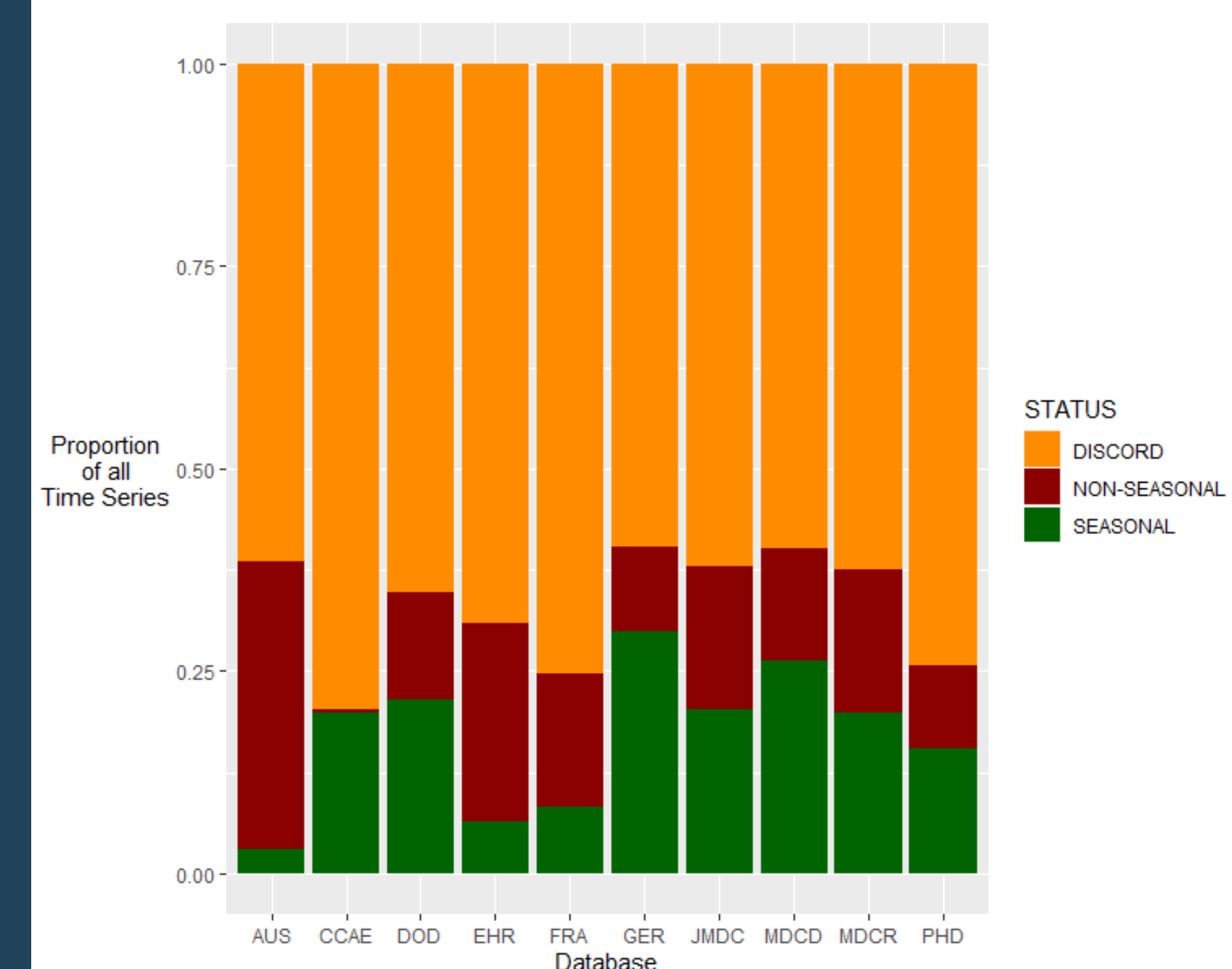
Methods for detecting seasonality in time series produce inconsistent results in observational healthcare data.



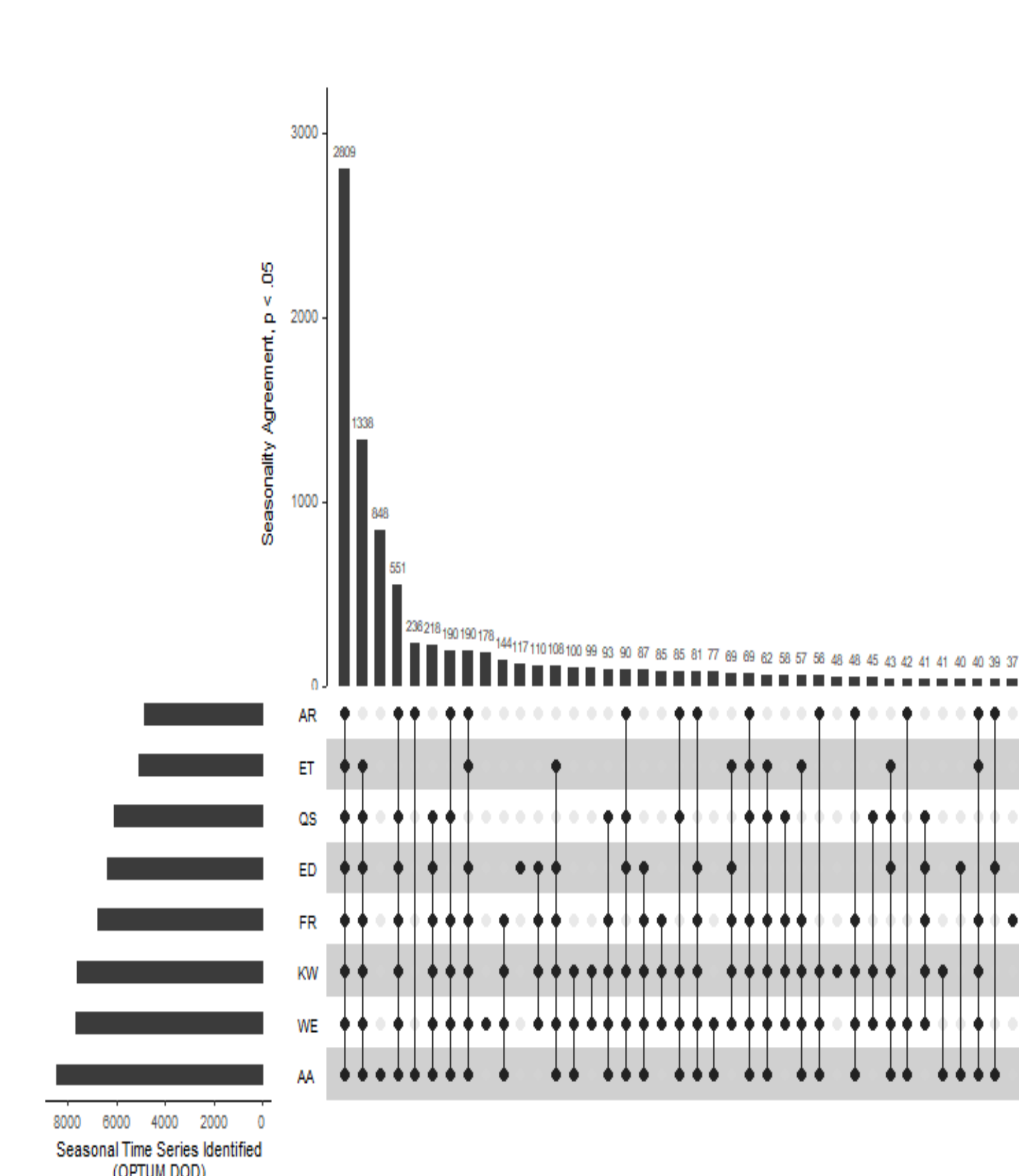
Results

- Using the 8 aforementioned statistical methods for determining seasonality, we evaluated 60,171 time series across 10 databases in the OMOP CDM format at three levels of significance. With more than 180,000 evaluations, we discovered substantial discord amongst the methods. This discord implies that these methods are not interchangeable and seasonality classification is highly dependent on the method chosen.

Concordance across all databases



UpsetR plot – concordance by method combination for OPTUM DOD, p < 0.05



Anthony Molinaro,
Frank DeFalco

