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

Anthony Molinaro, Frank DeFalco

Background

Seasonality classification is a well-known and important part of time series analysis. Understanding the seasonality of a biological event can contribute to an improved understanding of its causes and help guide appropriate responses. Observational data, however, are not comprised of biological events, but timestamped diagnosis codes the combination of which (along with additional requirements) are used as proxies for biological events. Given a visualization of an event of interest, the human eye can often detect repeatable patterns such as seasonality. Given large volumes of data, however, detection by eye is not feasible and automated methods must be employed. As there exist different methods for determining the seasonality of a time series, it is necessary to know if these methods are concordant with each other. In this study we seek to determine the concordance of these methods by applying them to time series derived from diagnosis codes in observational data.

Methods

We compared the 8 popular methods in Table 1 for determining the seasonality of a time series at three levels of significance (0.01, 0.05, and 0.1), against 10 databases in the OMOP CDM format.

Table 1. Method names, abbreviations, and brief descriptions.

METHOD NAME	ABB	BRIEF DESCRIPTION
Friedman's Test[1]	FR	Hypothesis test using ranks
Welch's Test[2]	WE	Hypothesis test using Welch's anova
Kruskal-Wallis Test[3]	KW	Hypothesis test using one way anova on ranks
Edwards' Test[4,6,11]	ED	Hypothesis test of a harmonic model of data fit using a Poisson glm
QS Test[5]	QS	Hypothesis test examining positive autocorrelation
ETS Hypothesis Test[7,8,9,10]	ETS	Hypothesis test to determine if the seasonal component is significant
Auto Arima Test[7,8,9,10]	AUTO_ARIMA	Test based on minimizing forecast errors across different models
ARIMA Hypothesis Test[7,8,9,10]	ARIMA	Hypothesis test to determine if the seasonal component is significant

As this study is concerned with contrasting methods of seasonality classification, it was most natural to create monthly time series representing how often condition concept identifiers occur in the data. The OHDSI package ACHILLES was used to aggregate the records associated with each condition concept identifier into monthly counts. The OHDSI package CASTOR was developed to transform these counts into proportions and create time series. The numerator of the proportion consists of the number of people (per thousand), with the condition concept identifier in each month, while the denominator consists of the number of people with an observation period spanning said month. For a concept to be eligible to be converted into a time series, we require at least four complete years (i.e., 12 months of counts each year) of data.

For each combination of database, method, significance level, and time series, we record the binary classification of seasonality. For each database and level of significance, we count the number of individual time series that are considered seasonal, compute the proportion seasonal, and compute concordance. We define concordance as unanimous agreement across all methods of binary seasonality classification for a given time series. Therefore, the methods are concordant only when they all classify a particular time series as either seasonal or non-seasonal. For the purposes of this study, the concern is not whether an individual method considers a given time series seasonal. Rather, the desired insight is whether all methods classify a given time series the same way. The concordance calculation is necessary because even identical proportions can hide disagreement. When two methods classify a similar proportion of time series as seasonal, it is useful to know whether the same individual time series were classified seasonal by both methods. This is impossible to determine by mere inspection of the proportion; two methods may classify the same number of completely different time series seasonal.

Data

The databases and the number of eligible time series evaluated are listed in Table 2, below.

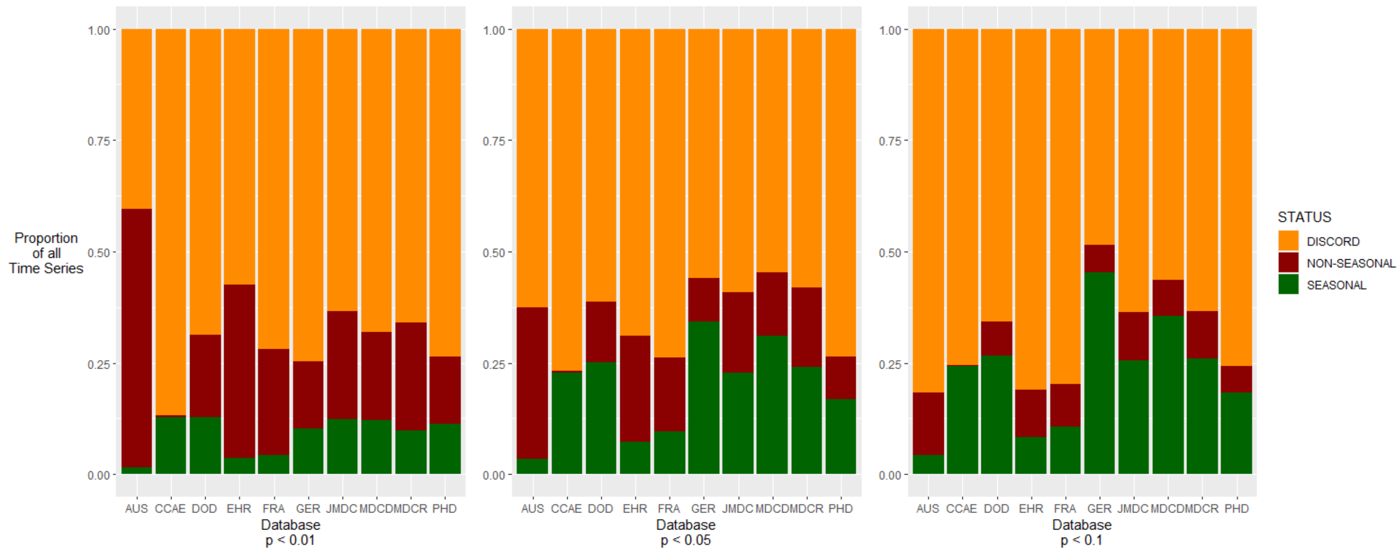
Table 2. Database names, abbreviations, and eligible time series count.

DATABASE	TIME SERIES
IQVIA Disease Analyzer – France (FRA)	896
IQVIA Disease Analyzer – Germany (GER)	3208
IQVIA Australian Longitudinal Patient Data (AUS)	408
Japan Medical Data Center (JMDC)	2956
IBM MarketScan Commercial Claims and Encounters (CCAE)	11051
IBM MarketScan Multi-State Medicaid (MDCD)	6478
IBM MarketScan Medicare Supplemental (MDCR)	6596
Optum Clinformatics Extended Data Mart - Date of Death (DOD)	11137
Optum Pan-Therapeutic Electronic Health Records (EHR)	12102
Premier Healthcare Database (PHD)	6635

Results

We evaluated 61,467 time series across 10 observational databases at three levels of significance (0.01, 0.05, and 0.1), totaling 184,401 evaluations. Figure 1 employs stacked bar charts to display the proportion of concordance across all databases and methods for each level of significance. Concordance is represented by the green and red bars. For $p < 0.01$, the range of concordance is 13.1% to 59.5%. For $p < 0.05$, the range of concordance is 23.1% to 45.2%. For $p < 0.1$, the range of unanimous agreement is 18.3% to 51.4%.

Figure 1. Stacked bar charts visualizing concordance by database across all significance levels.



Conclusion

Seasonality classification is a well-known and important part of time series analysis. Understanding the seasonality of a biological event can contribute to an improved understanding of its causes and help guide appropriate responses. However, given the existence of many ways of classifying an event of interest as seasonal and the presence of potentially thousands of events of interest, automated methods must be employed to take full advantage of large volumes of data. When relying on automated methods to assess the seasonality of time series derived from diagnosis codes in observational data, care should be taken since the results of this study indicate that methods are not interchangeable and seasonality determination is highly dependent on the method chosen. The high levels of discord that persist across all data sets and levels of significance imply that two researchers using different methods of binary seasonality classification are likely to observe different results.

References

1. Milton Friedman (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, Journal of the American Statistical Association, 32:200, 675-701, DOI: 10.1080/01621459.1937.10503522.
2. Welch, B. (1951). On the Comparison of Several Mean Values: An Alternative Approach. Biometrika, 38(3/4), 330-336. doi:10.2307/2332579
3. Kruskal, W., & Wallis, W. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, 47(260), 583-621. doi:10.2307/2280779.
4. Edwards JH (1961). The recognition and estimation of cyclic trends. Ann Hum Genet, 25:83-87. doi:10.1111/j.1469- 1809.1961.tb01501.x
5. Gomez, V. & Maravall, Agustin. (1997). Programs TRAMO and SEATS: instructions for the user. Mimeo, Banco de España.
6. Brookhart MA, Rothman KJ. Simple estimators of the intensity of seasonal occurrence. BMC Med Res Methodol. 2008;8:67. Published 2008 Oct 22. doi:10.1186/1471-2288-8-67
7. Hyndman R. (2010). https://stats.stackexchange.com/questions/1207/period-detection-of-a-generic-time- series/1214#1214
8. Hyndman R. (2012). https://robjhyndman.com/hyndsight/tscharacteristics

9. Hyndman, R. (2014). https://robjhyndman.com/hyndsight/detecting-seasonality/
10. Hyndman R. (2018). https://otexts.com/fpp2/seasonal-arima.html, https://otexts.com/fpp2/arima-ets.html, https://otexts.com/fpp2/arima-r.html
11. Weinstein RB, Schuemie MJ, Ryan PB, Stang PE. Seasonality in acute liver injury? Findings in two health care claims databases. *Drug Healthc Patient Saf*. 2016;8:39-48. Published 2016 Mar 31. doi:10.2147/DHPS.S95399