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

Anthony Molinaro (amolin19@its.jnj.com)
Janssen (United States)
Frank DeFalco
Janssen (United States)

Research Article

Keywords: Data Characterization, Observational Data, OMOP CDM, Seasonality, Time Series

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1044733/v1

Abstract

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. As there exist different methods for determining the seasonality of a time series, it is necessary to know if these methods exhibit concordance. 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 8 methods for determining the seasonality of a time series at three levels of significance (0.01, 0.05, and 0.1), against 10 observational health databases. We evaluated 61,467 time series at each level of significance, totaling 184,401 evaluations.

Results:

Methods of binary seasonality classification when applied to time series derived from diagnosis codes in observational health data produce inconsistent results. Across all databases and levels of significance, concordance ranged from 20.2% to 40.2%.

Conclusion:

The results indicate that researchers relying on automated methods to assess the seasonality of time series derived from diagnosis codes in observational data should be aware that the methods are not interchangeable. Seasonality determination is highly dependent on the method chosen.

Background

Seasonality classification is a well-known and important part of time series analysis. Events of interest (EOI) for which changes in frequency of occurrence follow a repeatable pattern based on calendar date are considered seasonal. Discovering whether an EOI is more likely to occur on a particular calendar date can contribute to an improved understanding of the EOI, its causes, and appropriate responses. Given a visualization of the frequency of occurrence of an EOI, the human eye can often determine whether a repeatable pattern, such as seasonality, exists. However, detection by eye is not feasible when working with large volumes of data containing thousands of potential EOI, therefore automated statistical methods must be employed. Necessarily, when relying on automated methods to discover true patterns, the existence of alternative methods and whether they are concordant should be known prior to investigation.

Observational data is patient level data comprised of prescription and health insurance claims, billing, and electronic health records. These data are assessed in various ways to determine whether they are appropriate for a given analysis. Healthcare researchers often attempt to assess the seasonality of an EOI by employing a variety of methods [1, 2, 3, 4, 5]. Given the existence of alternative methods, it is necessary to know if these methods exhibit concordance. 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

Data Sources

We used a total of 10 databases, each derived from either claims or electronic health records. We list the databases, abbreviations, and the number of time series evaluated. More detailed descriptions of the databases can be found in the appendix.

i. Premier Healthcare Database (PHD), 6635.
ii. Japan Medical Data Center (JMDC), 2956.
iii. Optum Electronic Health Records (EHR), 12102.
iv. IBM MarketScan® Commercial Claims and Encounters (CCAE), 11051.
v. IQVIA Disease Analyzer - France (FRA), 896.
vi. IQVIA Disease Analyzer – Germany (GER), 3208.
vii. IQVIA Australian Longitudinal Patient Data (AUS), 408.
viii. IBM MarketScan® Medicare Supplemental and Coordination of Benefits (MDCR), 6596.
ix. IBM MarketScan® Multi-State Medicaid (MDCD), 6478.
x. Optum Clinformatics Extended Data Mart - Date of Death (DOD), 11137.

Data Conversion and Time Series Creation

Each database had been previously converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [6]. The OMOP CDM organizes data into specific tables based on the type or domain of the data. The data used in this study comes from a table containing all condition occurrences where records are comprised of diagnosis codes and the corresponding dates when the codes were recorded in the data. Diagnosis codes in this table have been standardized to a unique identifier specified in the OMOP CDM vocabulary called a concept identifier.

As this study is concerned with contrasting methods of seasonality classification, it was most natural to create monthly time series objects representing how often these concept identifiers occur in the data. A tool called Achilles (Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems) [7], was used to aggregate the records associated with each condition concept

identifier into monthly counts. A package called CASTOR (Characterization and Analysis of Statistical Time series Of Real-world data) [8], 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.

Methods of binary seasonality classification

We evaluated 8 alternative methods for determining the seasonality of a time series at three levels of significance (0.01, 0.05, and 0.1), against 10 databases. For convenience, the methods evaluated are listed in Table 1. A more detailed description of the methods can be found in the appendix.

Table 1
Methods Summary

METHOD NAME	ABBREVIATION	BRIEF DESCRIPTION
Edwards' Test [9, 10, 11]	ED	Hypothesis test of a harmonic model of data fit using a Poisson generalized linear model. Seasonality is determined by evaluating the peaks and troughs of a sine curve fit to the observed time series.
Friedman's Test [12]	FR	Hypothesis test using a non-parametric approach for comparing samples within a population or from populations with identical medians. A rank-based approach is employed to test the hypothesis of no seasonality of the ranked months. Any linear trend in the data is removed prior to testing for seasonality.
ARIMA Hypothesis Test [13, 14, 18, 19, 23, 24]	AR	Hypothesis test to determine if the seasonal component is significant when compared to an identical ARIMA model without a seasonal component. Any linear trend in the data is removed prior to testing for seasonality.
QS Test [15]	QS	Hypothesis test to determine seasonality by examining the autocorrelation of seasonal lags. The observed time series is seasonal if positive autocorrelations at either lag 12 or 24 are significant. Any linear trend in the data is removed prior to testing for seasonality.
ETS Hypothesis Test [13, 14, 18, 19, 23, 24]	ET	Hypothesis test to determine if the seasonal component is significant when compared to an identical ETS model without a seasonal component. Any linear trend in the data is removed prior to testing for seasonality.
Kruskal-Wallis Test [16]	KW	Hypothesis test using a non-parametric approach to compare samples from a population. A rank-based approach is employed to test the hypothesis that the monthly data have the same mean. Any linear trend in the data is removed prior to testing for seasonality.
Welch's Test [17]	WE	Hypothesis test employing one-way ANOVA, but allowing for unequal variances amongst the groups of months. Seasonality is determined if hypothesis that the monthly means are identical is rejected. Any linear trend in the data is removed prior to testing for seasonality.
Auto ARIMA Test (5,6)	AA	Test based on minimizing forecast errors across different models. The observed time series is considered seasonal if the optimal ARIMA model chosen (the one that minimizes forecast error) includes a seasonal component. Any linear trend in the data is removed prior to testing for seasonality.

EXPERIMENT

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 for a given time series. Therefore, the methods are concordant 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 proportions are comprised of the same individual time series. This is impossible to determine by mere inspection of the proportion as an identical proportion may be had by classifying the same number of completely different time series.

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. Visualizations and tables were generated for each combination of database, method, and significance level. In an effort to provide a concise summary of the experiment, a subset of the results which is representative of the entire experiment will be presented.

Table 2 displays the proportion of time series classified seasonal by each method on all databases, for $p < 0.05$. Each row represents the results for all methods against a given database for $p < 0.05$. The method that classified the largest proportion of time series seasonal on a given database is highlighted in red. The method that classified the smallest proportion of time series seasonal in each database is highlighted blue. For instance, in the PHD database, the QS method classified 30.5% of the time series seasonal, while the AA method classified 79.2% seasonal. The method that classifies the least or greatest proportion of time series seasonal varies by database. KW, WE, AA, ED yielded the highest proportion in at least one database, while QS, ET, AR, and ED yielded the lowest proportion in at least one database. Each method varies substantially across databases; FR method identified 26% of time series in AUS as seasonal and 79.5% in GER.

Table 2
Proportion of time series classified seasonal, p < 0.05, blue indicates min, red indicates max

DATABASE	QS	FR	KW	WE	ET	AA	AR	ED
PHD	30.5%	45.7%	50.1%	47.7%	32.7%	79.2%	34.3%	62.8%
EHR	28.4%	49.9%	55.9%	47.3%	20.4%	43.3%	29%	59%
MDCR	46%	55.9%	64.1%	63.4%	36.3%	68.6%	42.3%	47.9%
MDCD	52.1%	66.7%	73.2%	69%	51.3%	70.3%	45.5%	56.1%
JMDC	47%	63.7%	68.4%	59.5%	33.5%	68.1%	47.5%	47.8%
GER	72.2%	79.5%	84.3%	83%	63.7%	75.1%	55.1%	49.8%
FRA	60.5%	50.9%	70.6%	71%	43%	67.2%	26.5%	36.9%
DOD	54.7%	61%	68.2%	69.1%	45.7%	76.2%	43.5%	57.5%
CCAE	52.9%	62.2%	68.4%	67.9%	43%	77.5%	42.1%	58.5%
AUS	23%	26%	35.3%	38.5%	9.6%	33.1%	17.6%	33.1%

Figure 1 displays the proportion of concordance across all databases, for all methods and levels of significance. Concordance is represented by the green and red bars. The range of concordance is 20.2–40.2%.

For further exploration into the behavior of the individual methods, we provide the following statistics, figures, and tables from OPTUM DOD, p < 0.05. On this database, the methods exhibit concordance for 4,307 time series; classifying 2,809 as seasonal and 1,498 as non-seasonal. The mean and max variance for the 2,809 time series classified seasonal are 0.031 and 18.4, respectively. The mean and max variance for the 1,498 time series classified non-seasonal are 0.000014 and 0.019265, respectively.

Figure 2 is an UpsetR plot that visualizes 40 different combinations of seasonality classification on OPTUM DOD, p < 0.05. To the left of the method names is a bar chart that shows the number of time series classified seasonal by each method. The dots with the lines through them indicate which method participated in each combination. Reading from left to right, we explain the first four combinations of methods. The first combination indicates that there were 2,809 time series for which all methods agreed were seasonal. The second combination indicates that 1,338 time series were classified seasonal by all methods except AR. The third combination indicates that there were 848 time series that only the AA method classified as seasonal. The fourth combination indicates that there were 551 time series classified seasonal by all methods except ET.

Figure 3 is a 3x3 plot of nine time series and their binary seasonality classification by each method on OPTUM DOD, p < 0.05. Atop each time series is the abbreviation for each method. A color-coding scheme was used to indicate whether a method classified the given time series as seasonal (green) or non-

seasonal (red), respectively. As per Table 1, any linear trend that appears in the original time series is removed prior to testing for seasonality. Beneath each time series is the corresponding concept identifier, name, and two different counts. The value for N represents the number of times the specified (green-red) combination occurred, while the value for M represents the number of times a similar combination occurred. For instance, AR, FR, ET, and AA all classified “Disorganized schizophrenia” (the center plot) as non-seasonal. N = 10 implies that there were 10 time series classified as non-seasonal by this specific combination of methods. In this case, four of the methods classified the time series as seasonal, while four did not. M = 602 implies that there were 602 time series for which (any combination of) four methods classified as seasonal while four did not. In Figure 3, the methods exhibit concordance for only two time series: Frostbite of foot and Large cell anaplastic lymphoma (top-left corner and bottom-right corner, respectively).

Discussion

The purpose of this study was to determine whether there exists concordance among different methods of binary seasonality classification when applied to time series derived from diagnosis codes in observational data. The results of this study, as shown in Figure 1, indicate the methods are generally inconsistent with each other, with discordance observed in 60–80% of time series across 10 populations. As Table 2 reveals, the methods exhibit variation both across databases and within databases, implying that the source of the variation is not the data, but the methods themselves. Ultimately, the source of discord stems from the different ways in which the methods assess seasonality. While there do exist similarities, each method focuses on a different aspect of a time series to assess seasonality (Table 1). For instance, half the methods fit a time series with a hypothetical model and test the model for seasonality, while the other half test different aspects of a time series directly, without using a hypothesized model. To take the discussion further and generalize where we can, we make distinctions between types of concordance and types of peaks. Regarding concordance, we define “positive concordance” to be unanimous agreement among the methods that a time series is seasonal, while “negative concordance” to be unanimous agreement that a time series is non-seasonal. Therefore, for a given time series, the methods are discordant when there is neither positive concordance nor negative concordance. Regarding peaks, we say that peaks are “persistent” if they occur year after year, and they are “consistent” if they occur in the same month. We make this distinction because peaks relate to important aspects of time series analysis relevant to seasonality; specifically, variation and autocorrelation. Peaks can, of course, come in different sizes. Time series with large peaks suggest greater variation than those with small peaks. Persistent peaks (be they small or large) suggest the possibility of underlying cyclical behavior in the time series. Consistent peaks, to the extent that they are consistent, indicate autocorrelation in the time series. We’ll use Figures 2 and 3 to navigate the remainder of the discussion. For the sake of brevity, when discussing the individual time series in Figure 3, reading from top-left to bottom-right, we’ll refer to them as Fig. 3.ts1, Fig. 3.ts2, …, Fig. 3.ts9.

From Fig. 3.ts1 (N = 2,809) and Fig. 3.ts9 (N = 1,498), we learn that the methods exhibit concordance only 4,307/11,137 = 38.7% of the time. Figure 2 provides valuable insight into the extent of discord among the

methods. Of the 40 unique combinations, we observe that some combinations occur more frequently than others and this is due to similarities in the testing procedure (Table 1). For instance, methods that group time series data by month and test for differences among the groups are assessing seasonality differently than methods that fit a hypothetical model and then determine seasonality by minimizing forecast error. Acknowledging the differences in how the methods assess seasonality is important not only for understanding the amount of observed discord, but in recognizing that these differences indicate a disagreement with regards to how seasonality is defined. Indeed, if the methods were highly concordant despite their contrasting approaches, we would have to concede that the contrasting approaches are ultimately just different ways of expressing the same aspect of a time series. This can be observed more clearly by exploring Figure 3. In Fig. 3.ts1, …, Fig. 3.ts4 we observe time series that to the human eye seem seasonal and very similar. Identifying such time series as seasonal is a very old idea in time series analysis, with Beveridge [26] and Yule [27] employing harmonic functions to model time series with cyclical behavior. However, despite an obvious cyclical pattern and visual similarities, Fig. 3.ts2, Fig. 3.ts3, and Fig. 3.ts4, all exhibit discord. The reason being, except for the ED method, the methods are not testing for seasonality by fitting the data with harmonic functions. Thus, the different methods of seasonality assessment ultimately result in different definitions of seasonality.

As we've mentioned previously, the behavior of peaks plays an important role in concordance. We'll use Figure 3 further to explore the relationship between peaks, variation, and discord, and provide general principles as to when a method would be more likely to classify a time series as seasonal rather than non-seasonal.

Since each method assesses seasonality differently, positive concordance is only achieved when multiple conditions are simultaneously present. Persistent and consistent peaks are most important for ED, AA, AR, and ET. Peaks will result in a seasonal classification by ED, so long as there exists a sufficient difference between the peaks and troughs in the data. However, even with persistent and consistent peaks, variation (particularly among the peaks) over time can lead to a non-seasonal classification by AA, AR, or ET (Fig. 3.ts2, Fig. 3.ts3, and Fig. 3.ts4). Indeed, we have confirmed experimentally that we can achieve positive concordance for the time series in Fig. 3.ts2, Fig. 3.ts3, and Fig. 3.ts4, by removing the data prior to 2016. Since time series with persistent and consistent peaks will have high correlation between seasonal lags, they will be classified seasonal by QS. For FR, KW, and WE, most important is variation. In the absence of the prominent peaks we see in Fig. 3.ts1, …, Fig. 3.ts4, sufficient variation in the time series data can lead FR, KW, and WE to a seasonal classification (Fig. 3.ts6). Therefore, with regards to positive concordance we see tension among the methods in that variation may cause some methods to classify seemingly seasonal time series as non-seasonal (Fig. 3.ts2, Fig. 3.ts3, and Fig. 3.ts4) and seemingly non-seasonal time series as seasonal (Fig. 3.ts5, …, Fig. 3.ts8).

The relationship between negative concordance and variation is more straightforward. The time series in Fig. 3.ts5, …, Fig. 3.ts9 are similar in that one cannot determine the results of the methods by visual inspection alone (recall that any linear trend in each of the original series have been removed prior to method application). Given the similarity of the time series in Fig. 3.ts5, …, Fig. 3.ts9, it's reasonable to

wonder why they all do not exhibit negative concordance. Ultimately, time series that are constant or stationary around a constant mean with minimal variation will result in negative concordance among the methods. However, a time series with both large peaks and variation will exhibit negative concordance if there is no monthly or yearly autocorrelation (for instance, a time series generated from $N(\mu,\sigma^2)$). As was noted in the Results section, the 1,498 time series for which the methods exhibit negative concordance report a mean variance of 0 to four decimal places.

We've explained general scenarios in which we can expect negative and positive concordance, but further generalization in more difficult. As Figure 3 reveals, there are thousands of different combinations of discord (M = 2,168, …, 1,267) for each time series, making it difficult to predict which particular combination of discord to expect based on visual inspection of the time series alone. However, an immediate consequence of this study is that researchers using different methods are implicitly defining seasonality differently. Given the discordance between the methods, researchers relying on different methods are likely to encounter different results, thus leading to conflicting understanding of the seasonality of a time series.

Finally, we note that the study and evaluation of methods was limited to 10 observational databases and eight methods of binary seasonality classification. Different results may have been observed by modifying one or more of the following design choices:

- Construction of time series
- Number and choice of databases
- Number and choice of methods of binary seasonality classification

Conclusion

The results of this study indicate that the determination of the seasonality of a time series is highly dependent on the method chosen. The methods are not interchangeable and lead to vastly different results between and within databases. Consequently, researchers investigating seasonality with these methods must be aware that their results are not generalizable to other methods. Researchers investigating seasonality with these methods should be aware that their choice of method implies how they define seasonality in their study. Consequently, the method chosen should be listed as a limitation of a study. The results of this study indicate that while seasonality may be intuitively understood, it is not well defined with regards to automated statistical tests.

List Of Abbreviations

Methods: Auto ARIMA Test (AA), ARIMA Hypothesis Test (AR), Edwards' Test (ED), ETS Hypothesis Test (ET), Friedman's Test (FR), Kruskal-Wallis Test (KW), QS Test (QS), Welch's Test (WE).

Databases: IBM MarketScan® Medicare Supplemental and Coordination of Benefits (MDCR), IBM MarketScan® Multi-State Medicaid (MDCD), IBM MarketScan® Commercial Claims and Encounters (CCAE),

IQVIA Disease Analyzer - France (FRA), IQVIA Disease Analyzer – Germany (GER), IQVIA Australian Longitudinal Patient Data (AUS), Japan Medical Data Center (JMDC), Optum Electronic Health Records (EHR), Optum Clinformatics Extended Data Mart - Date of Death (DOD), Premier Healthcare Database (PHD).

Declarations

Ethics approval and consent to participate: Not Applicable.

No human participants, human material, or human data was used in this study.

Consent for publication: Not Applicable.

Availability of data and materials:

- https://products.premierinc.com/applied-sciences
- https://www.ibm.com/products/marketscan-research-databases
- https://www.iqvia.com/
- https://www.jmdc.co.jp/en/jmdc-claims-database/
- https://www.optum.com/business/about/data-analytics-technology.html
- https://github.com/OHDSI/Achilles
- https://github.com/OHDSI/Castor
- https://github.com/OHDSI/CommonDataModel
- https://ohdsi.github.io/TheBookOfOhdsi/

The databases used in this study are all commercial databases licensed from IBM, Optum, Iqvia, and JMDC, respectively.

Competing interests:

Both authors are full time employees of Janssen Research and Development, a unit of Johnson and Johnson.

The work on this study was part of their employment. They also hold pension rights from the company and own stock and stock options.

Funding: Not Applicable.

The work on this study was part of the authors' employment at Janssen Research and Development. No additional funding was supplied beyond compensation as Janssen employees.

Authors' contributions

AM and FD designed and developed the experiment and leveraged tools developed by themselves and the OHDSI (Observational Health Data Analytics) community. Both authors contributed to the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Jesse Berlin, Mitchell Conover, and Martijn Schumie of Janssen Research and Development for their help in developing the experiment.

The authors would like to especially acknowledge Patrick Ryan of Janssen Research and Development for his substantial contributions to the final manuscript.

References

1. Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. PLoS Pathog. 2018;14(11):e1007327. Published 2018 Nov 8. doi:10.1371/journal.ppat.1007327
2. Yoon JY, Cha JM, Kim HI, Kwak MS. Seasonal variation of peptic ulcer disease, peptic ulcer bleeding, and acute pancreatitis: A nationwide population-based study using a common data model. Medicine (Baltimore). 2021;100(21):e25820. doi:10.1097/MD.0000000000025820
3. Fisman DN. Seasonality of infectious diseases. Annu Rev Public Health. 2007;28:127-43. doi: 10.1146/annurev.publhealth.28.021406.144128. PMID: 17222079.
4. Fisman D. Seasonality of viral infections: mechanisms and unknowns. Clin Microbiol Infect. 2012 Oct;18(10):946–54. doi: 10.1111/j.1469-0691.2012.03968.x. Epub 2012 Jul 20. PMID: 22817528.
5. Ramanathan K, Thenmozhi M, George S, Anandan S, Veeraraghavan B, Naumova EN, Jeyaseelan L. Assessing Seasonality Variation with Harmonic Regression: Accommodations for Sharp Peaks. Int J Environ Res Public Health. 2020 Feb 18;17(4):1318. doi: 10.3390/ijerph17041318. PMID: 32085630; PMCID: PMC7068504.
6. OMOP Common Data Model (https://ohdsi.github.io/CommonDataModel/) Accessed 20 Oct 2021.
7. Achilles (https://github.com/OHDSI/Achilles) Accessed 27 June 2019.
8. Castor (https://github.com/OHDSI/Castor) Accessed 2 Oct 2020.
9. Edwards JH. The recognition and estimation of cyclic trends. Ann Hum Genet, 1961;25:83–87. doi:10.1111/j.1469-1809.1961.tb01501.x
10. 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

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
12. Milton Friedman. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, Journal of the American Statistical Association, 1937;32:200, 675–701, DOI: 10.1080/01621459.1937.10503522
13. Hyndman, R. Detecting Seasonality. https://robjhyndman.com/hyndsight/detecting-seasonality/ (2014). Accessed 27 June 2019.
14. Hyndman R, Athanasopoulos G. Forecasting: Principles and Practice. Online Edition. https://otexts.com/fpp2/, https://otexts.com/fpp2/seasonal-arima.html, https://otexts.com/fpp2/arima-ets.html, https://otexts.com/fpp2/arima-r.html, https://otexts.com/fpp2/ets.html, https://otexts.com/fpp2/estimation-and-model-selection.html (2018). Accessed 27 June 2019.
15. Gomez, V. & Maravall, Agustin. Programs TRAMO and SEATS: instructions for the user. Mimeo, Banco de España. (1997).
16. Kruskal, W., & Wallis, W. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, 1952;47(260), 583–621. doi:10.2307/2280779
17. Welch, B. On the Comparison of Several Mean Values: An Alternative Approach. Biometrika, 1951;38(3/4), 330–336. doi:10.2307/2332579
18. Hyndman R. Period detection of a generic time series. https://stats.stackexchange.com/questions/1207/period-detection-of-a-generic-time-series/1214#1214 (2010). Accessed 27 June 2019.
19. Hyndman R. Measuring time series characteristics. https://robjhyndman.com/hyndsight/tscharacteristics (2012). Accessed 27 June 2019.
20. Shumway R. and D. Stoffer. Time Series Analysis and Its Applications With R Examples (3rd Ed), Springer, 2011.
21. Webel, K. and D. Ollech. An overall seasonality test based on recursive feature elimination in conditional random forests. Proceedings of the 5th International Conference on Time Series and Forecasting, 20–31, 2018.
22. Seastests (https://cran.r-project.org/web/packages/seastests/index.html): Seasonality Tests - An overall test for seasonality of a given time series in addition to a set of single seasonality tests as used in Ollech and Webel (forthcoming): An overall seasonality test. Bundesbank Discussion Paper.
23. Hyndman, R. J., and Khandakar, Y. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 2008; 27(1), 1–22. https://doi.org/10.18637/jss.v027.i03
24. Forecast (https://cran.r-project.org/web/packages/forecast/index.html): Forecasting Functions for Time Series and Linear Models. Methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.

25. Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann. Math. Statist. 9, 1938, no. 1, 60–62. doi:10.1214/aoms/1177732360.
26. Beveridge W.H. Wheat Prices and Rainfall in Western Europe. Journal of the Royal Statistical Society, Vol. 85, No. 3, 1922, pp. 412–475
27. Yule G.U. Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. Journal of the Royal Statistical Society, Vol. 89, No. 1, 1926, pp. 1–63

Figures

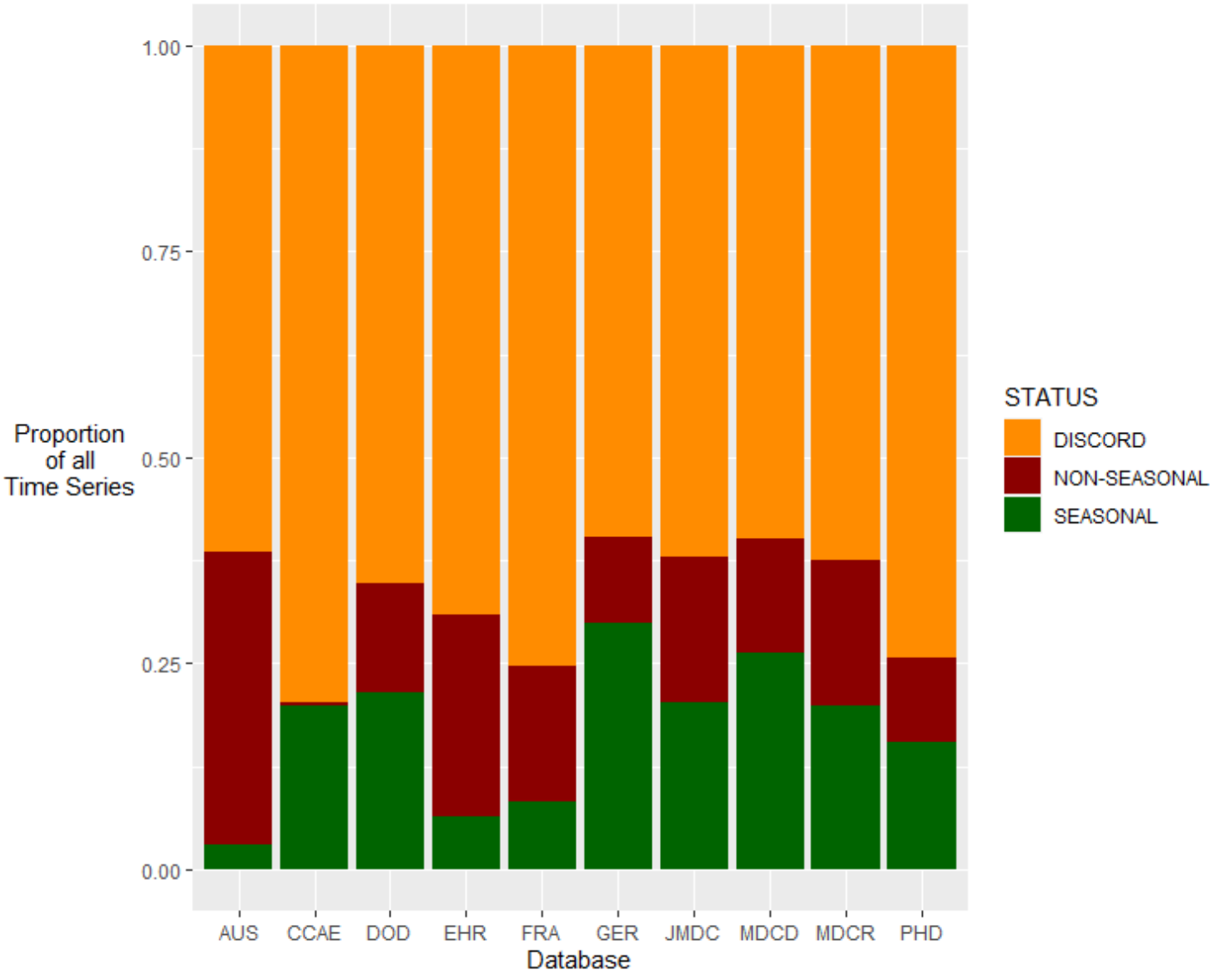

Figure 1

Stacked bar chart visualizing concordance by database across all significance levels.

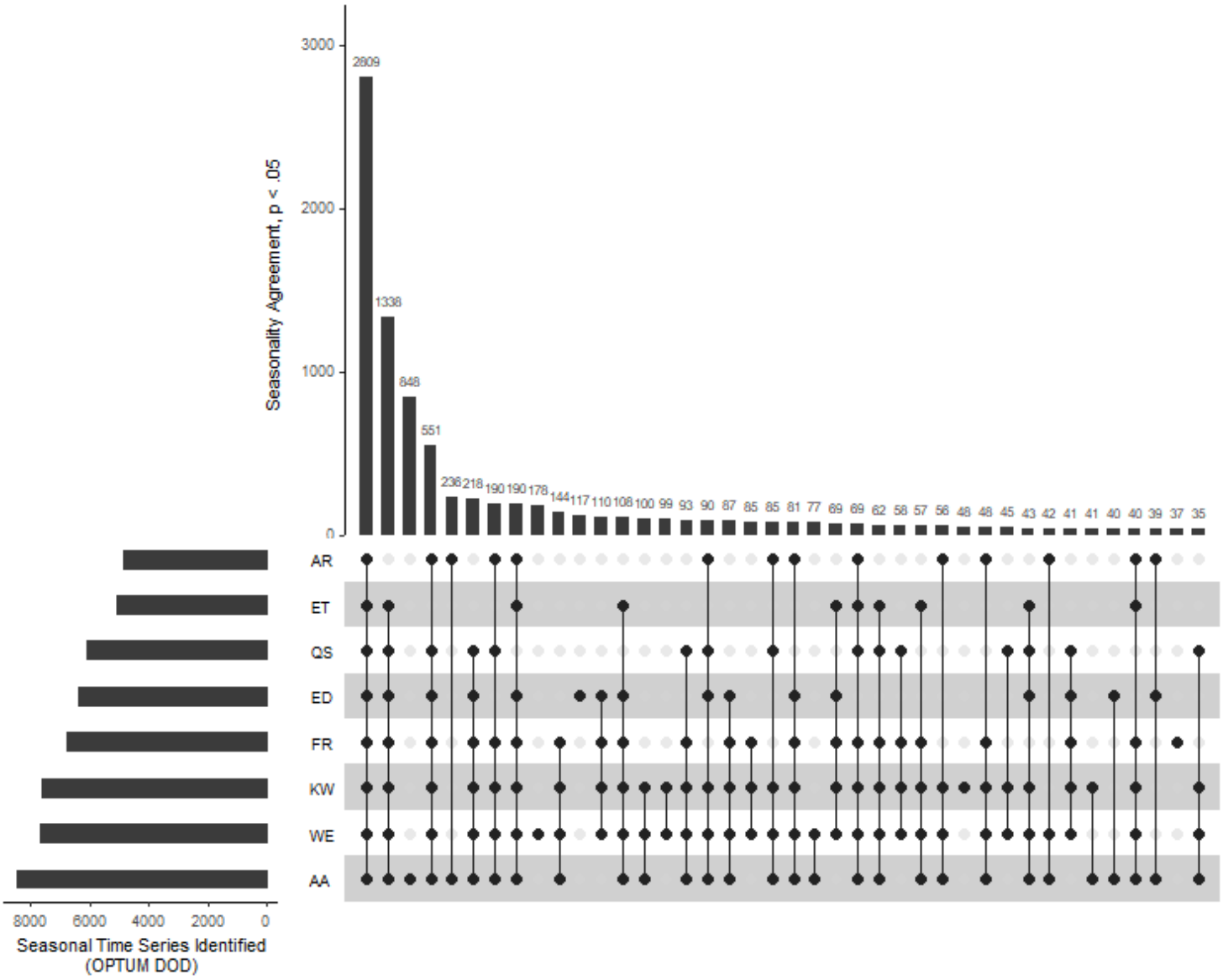

Figure 2

UpSetR plot visualizing 40 different method combinations of seasonality classification for OPTUM DOD, $p < 0.05$.

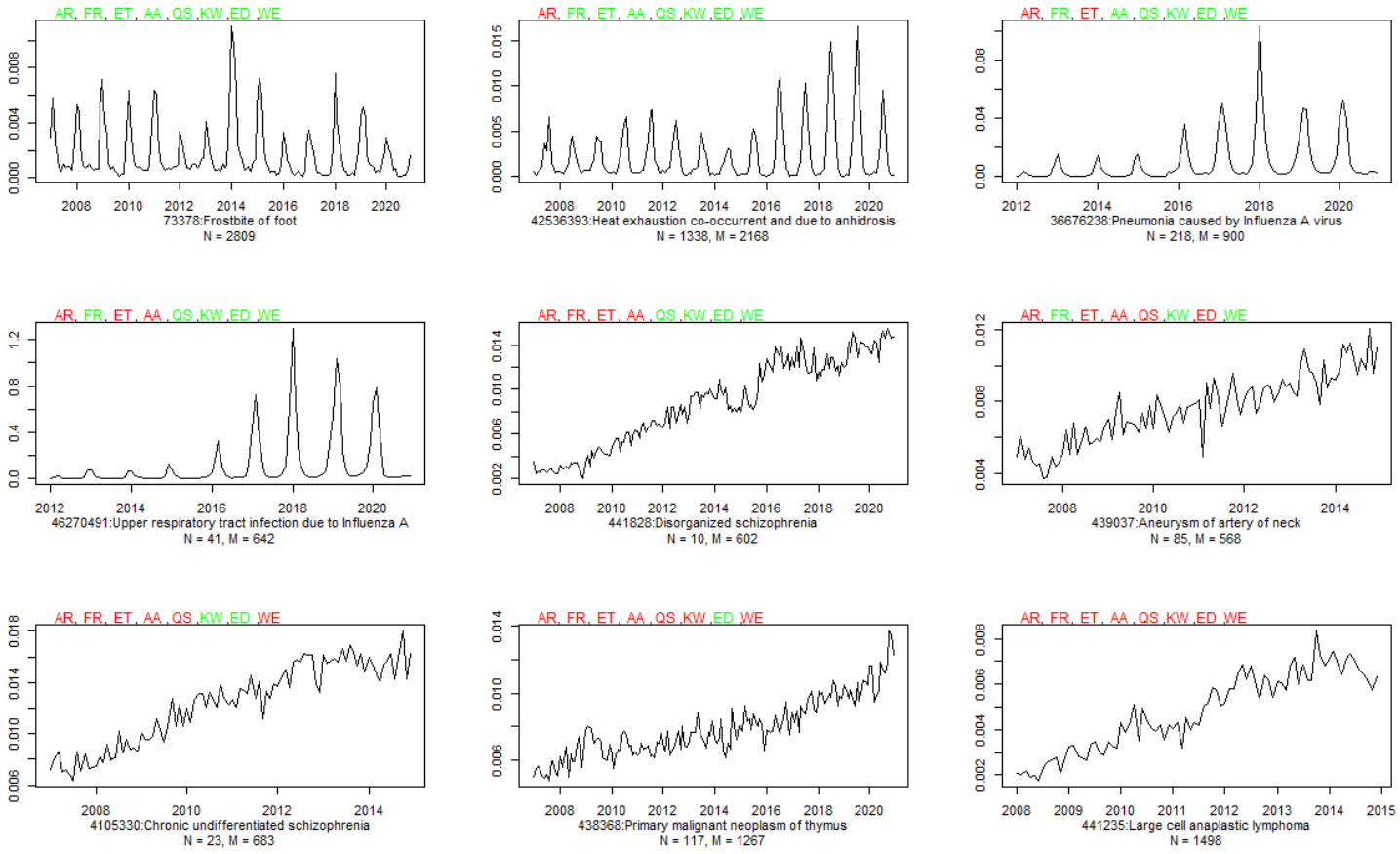

Figure 3

Nine time series from OPTUM DOD and their binary classification by each method, $p < 0.05$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx
- AdditionalFile2.docx