Name:	Yohan Sumathipala
Affiliation:	National Institutes of Health – National Library of Medicine (NIH-NLM)
Email:	sumathipalaya@mail.nlm.nih.gov; yohan123@gmail.com
Presentation type (s):	**Poster**

Classification of temporal trends in healthcare research databases in support of data quality and hypothesis generation

Yohan A. Sumathipala, Vojtech Huser, MD, PhD
National Library of Medicine, Bethesda, MD, USA

Abstract

Healthcare systems generate massive amounts of data on the occurrence of medical events, which include medical procedures, prescription drugs, and patient diagnoses. Understanding extract insights to improve patient safety and care outcomes is an important goal of modern medicine. Extracting such insights requires both high quality data and a clear hypothesis to test against the data. In this research, we developed and applied a set of statistical methods to verify the quality of large EHR datasets and to assist in hypothesis generation. These methods identify seasonality at the monthly and daily level. From a large database of claims data, we identified seasonality in 7,761 procedures, 1,351 drug ingredients, and 10,680 diagnoses. In this work, we present our methodologies and surprising seasonal events they elucidated.

Introduction

Healthcare systems generate massive amount of data through electronic health record (EHR) and healthcare claims. Leveraging this data to improve patient safety, outcomes and further the science of care delivery is widely recognized a tenet and goal of 21st century healthcare. To help with data quality assessment and hypothesis generation we analyzed temporal trends in medical events such as medications, procedures, diagnoses, and lab results.

Methods

We developed methodologies to extract 13 different features from time-series data to classify trends. These features are: period, trend, seasonality, autocorrelation, non-linearity, skewness, kurtosis, Hurst, Lyapunov, seasonally adjusted autocorrelation, seasonally adjusted non-linearity, seasonally adjusted skewness, and seasonally adjusted kurtosis. We use these features to identify and characterize seasonality by week and by month in data that follows the Observational Medical Outcomes Partnership Common Data Model (CDM). Using these statistical methods, we analyzed 7,761 procedures, 1,351drug ingredients and 10,680 diagnoses, using claims data.

Results

Our results showed the following expected seasonal events: preventive care (peaked in August); influenza vaccines (fall); and pulmonary conditions (winter). Surprising seasonal events included adenoidectomies (spring) and visual field exams (August); fluorouracil, atenolol and dextromethorphan (January and winter); and acute pyelonephritis (August) and concussions (September-October). Analysis of weekly seasonality revealed knee arthroplasties are 17% more common on Mondays ($p < 0.001$) and, surprisingly, vasectomies were 30% higher on Fridays ($p < 0.001$).

Conclusions

Detection of seasonal trends can be used to assess data quality, complimenting conventional rule-based approaches. The R software package developed in this research can be applied to explore seasonality in any CDM dataset to discover geographic and international differences; we expect Southern Hemisphere nations to exhibit seasonality patterns different from the U.S.