Detecting effects of holidays and seasonality on female breast cancer incidence using central cancer registry data

Qiao Wang, MA[1,2]; **Chester L. Schmaltz**, PhD[1,3]; **Jeannette Jackson-Thompson**, PhD, MSPH[1,3,4]; **Yilin Yoshida**, PhD, MPH[1,3]; **Dongchu Sun**, PhD[2]

[1]*Missouri Cancer Registry and Research Center (MCR-ARC);* [2]*University of Missouri-Columbia (MU), College of Arts & Sciences, Department of Statistics;*
[3]*MU School of Medicine, Department of Health Management & Informatics;* [4]*MU Informatics Institute, Columbia, Missouri*

1. Background

- Effects of holidays and seasonality have been widely studied in fields such as economics and neuroscience.
- Few efforts have been made to perform time series analysis on the cancer registry data, which is a great source of time series analysis due to its long time period.

2. Objective

Apply time series analysis techniques to daily and monthly incidence of invasive female breast cancer (FBC) to quantify holiday effects and seasonality.

3. Data Overview

- We used Missouri Cancer Registry data. It included complete data (YYYYMMDD) on incident cases of invasive FBC from 1996 to 2015 in Missouri (N=84,792 cases).
- Daily incidence data included 7,305 days and 1,114 days had zero counts (See Figure 1).
- Figure 2 presented times series plot of daily incidence data in 2015 as an example to illustrate the structure of daily incidence data and the fact that weekends had much lower incidences than weekdays. The coral points stand for significant holiday effects (See Result).
- Figure 3 showed monthly incidence data with 20 subseries (years) and 240 months in total. The blue line represents the average incidence of the corresponding month over 20 years.

Figure 1: Histogram of daily incidence for invasive FBC, 1996-2015

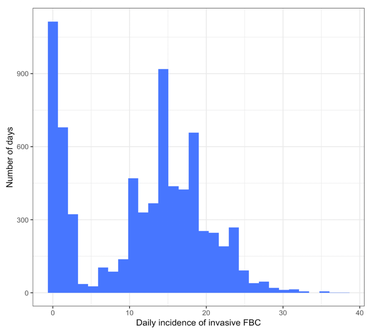


Figure 2: Daily incidence of invasive FBC for weekdays and weekends in 2015

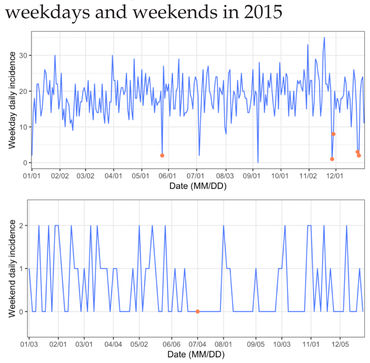
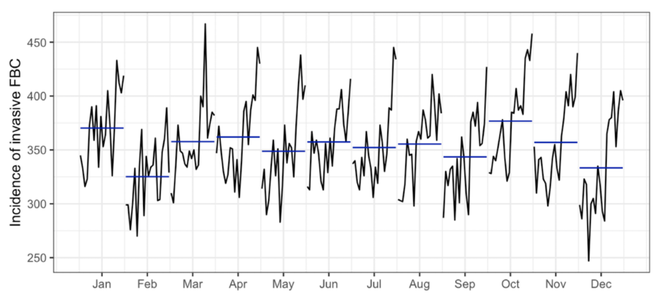


Figure 3: Seasonal subseries plot for incidence of invasive FBC 1996-2015



4 Methods

Daily Incidence Data

- Zero-inflated Poisson regression model with lagged variables was used to adjust for **excess zero** counts in the data. Covariates included a **linear trend**, **holiday** effects, **weekend** effects and **seasonality**.
- "Holidays" included New Year's Day, Birthday of Martin Luther King Jr., Good Friday, Memorial Day, Independence Day, Labor Day, Thanksgiving Day, Black Friday, Christmas Eve and Christmas Day.

Monthly Incidence Data

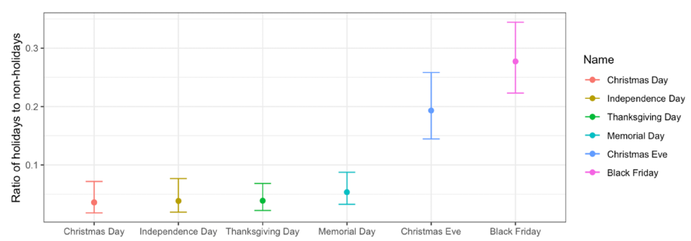
- The main goal was to **decompose** the daily incidence into **trend**, **seasonality** and irregular (white noise), which we exptected to be random. A seasonal-trend decomposition procedure based on Locally Weighted Scatterplot Smoothing (LOWESS) was used to obtain the structure of trend and seasonality.

5. Results

Daily Incidence Data

- Six out of ten holidays had a significant influence on the incidence based on AIC and p-value ($< 10^{-6}$). Incidence of invasive FBC on **holidays** was **lower** than non-holidays. Among holidays, Christmas Day, Independence Day and Thanksgiving Day had the lowest incidence (See Figure 4).
- **Weekends** had a great **negative** effect on the incidence of invasive FBC. The incidence ratio of weekend to weekday was 0.1022 with CI [0.0954, 0.1095]. The probability to get a zero incidence on the weekend was 37.4 % with CI [37.07%, 37.76%]. This agreed with time series plot for weekdays and weekends in Figure 2.
- **Seasonality** of one week, two weeks, one month and one year were **detected**. We also noticed that seasonality was relatively small compared with weekend and holiday effects.
- **On average**, we expected **0.04**% **more** incidence per year.

Figure 4: Incidence ratio of holidays to non-holidays



Monthly Incidence Data

- **Overall**, there was an **increasing** trend with fluctuations over 1996-2015 (See Figure 5).
- Figure 6 showed the detailed seasonal structure of monthly data. More incidences were expected in Jan., Mar., Apr., Jun., Aug., Oct. and Nov.
- A seasonality had been spotted with a **valley** in **Feb.** and a **peak** in **Oct.** Notice that Oct. is the **awareness month** for breast cancer (See Figure 6). This was consistent with Figure 2 results.

5. Results (Continued)

Figure 5: Decomposition of monthly incidence data

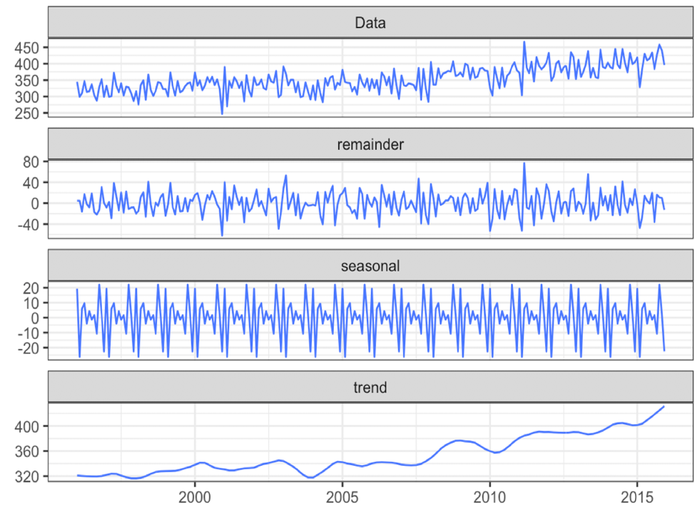
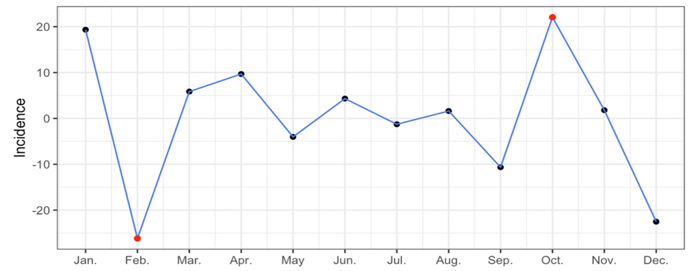


Figure 6: Seasonal structure of monthly incidence data



6. Conclusions

- Both daily and monthly incidence data indicated seasonality.
- Holidays and weekends had a great negative impact on the incidence of invasive FBC.

7. Discussion

- Daily data had the advantage of studying holidays effects and short term seasonality such as different weekly effects and weekend effects.
- The long term seasonality such as annual effects and different monthly effects could be easily masked by short term seasonalities. In this case, monthly incidence data gave us more power to study the long term seasonality.
- When seasonality is indicated in the data, identifying the seasonality stucture is beneficial for two reasons: (1) Help authorities to make decisions regarding seasonal patterns; (2) Provide better predictions for the variable of interest.

Acknowledgement: MCR-ARC core activities are supported in part by a cooperative agreement between the Centers for Disease Control and Prevention (CDC) and the Missouri Department of Health and Senior Services (DHSS) (NU58DP006299-01/02) and a Surveillance Contract between DHSS and the University of Missouri.
We would like to thank the MCR-ARC Quality Assurance staff, and the staff of facilities throughout Missouri and other states' central cancer registries for their dedication and desire for continuous quality improvement and submitting their reportable cases to MCR-ARC.

Contact: Qiao Wang (qwxv2@mail.missouri.edu).