Differences in holiday and seasonality effects on female breast cancer incidence in Missouri by age, stage and rural vs. urban

Qiao Wang, MA[1,2]; **Chester L. Schmaltz**, PhD[1,3]; **Jeannette Jackson-Thompson**, PhD, MSPH[1,3,4]; **Yilin Yoshida**, PhD, MPH[3,5]; **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;* [5]*Tulane University, School of Medicine, Section of Endocrinology, Department of Medicine.*

1. Background

- Few efforts have been made to explore time series patterns (seasonality and holiday effects) on the cancer registry data, which is a great source of time series analysis due to its long time period and high quality.

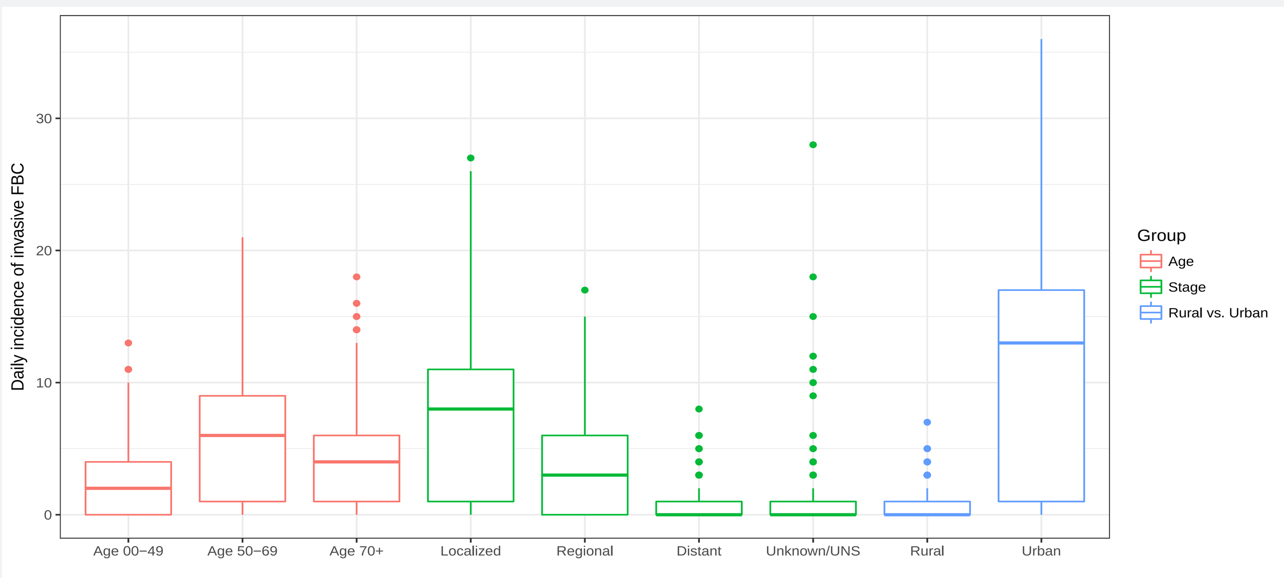
2. Objective

- Perform 9 different univariate time series analysis regarding daily incidence of invasive female breast cancer (FBC) to examine individual holiday effect and seasonality respectively.

3. Data overview

- We used Missouri Cancer Registry data, which included complete diagnosis date (YYYYMMDD) on incident cases of invasive FBC from 1996 to 2015 in Missouri (N=84,792 cases).
- 9 invasive FBC univariate daily time series were
  - Two geographic unit groups: rural and urban;
  - Three age groups: 00-49, 50-69 and 70+;
  - Four stage groups: localized, regional, distant and unknown/unstaged (UNS).
- Each daily time series included 7,305 days. Urban had the least zero counts (1,155 days) and UNS had the most (5,188 days).

Figure 1: Daily incidence distribution by age, stage and rural vs. urban



4 Methods

- For each daily time series, autoregressive conditional Poisson (ACP) and zero-inflated Poisson (ZIP) regression models were used, where ZIP was used to adjust for **excess zero** counts in the data if necessary.
- Covariates included a **linear trend**, **holiday** effects, **weekend** effects and **seasonality**.
  - Eight holidays were included: New Year's Day, Good Friday, Memorial Day, Independence Day, Thanksgiving Day, Black Friday, Christmas Eve and Christmas Day;
  - Five seasonal patterns were included: one day, one week, semi-month, one month and one year.
- Model selection was based on p-value (if available), Akaike information criterion (AIC) and auto-correlation function (ACF) plot for residuals.

5. Results

If the estimated relative risk (incidence ratio) for a covariate is less than 1, there is a negative impact. A smaller incidence ratio indicates a larger negative impact.

- For all 9 univariate time series, **weekends** had the greatest **negative** effect.
  - The smallest incidence ratio of weekend to weekday is 0.0538 (age 50-69) and the largest is 0.6556 (unknown/UNS).
- A **positive** linear trend was detected, which implied that we expected **more** incidence per year except age 00-49 (no increase/decrease), rural (decrease) and UNS (decrease).
- There were **less** incident cases on **holidays** and differences varied both within and across age, stage and rural vs. urban (See Tables 1-3, the smallest incidence ratio is highlighted in yellow. Only significant holidays were reported.)
  - "Thanksgiving Day" had the least incident cases for more than half of 9 groups.
  - "Good Friday" had the least impact on incident cases for more than half of 9 groups compared with other holidays.
  - "New Year's Day" for age 70+ showed a positive effect.
  - Unknown/UNS showed no holiday effects except one positive holiday effect on "New Year's Day".
- Seasonal patterns (See Table 4) existed and relatively small compared with other covariates.
  - Rural showed no seasonal patterns while the rest of 9 groups showed at least one type of seasonal patterns.
  - Localized had all types of seasonal patterns.
  - "One year" was the most common seasonal pattern.
- "Rural" and "unknown/UNS" groups required further investigation due to their high percentage of zero counts and small non-zero incidence counts (See Figure 1). For example, on Independence Day, rural had no incidence over 20 years.

Table 1: Incidence ratio of holidays to non-holidays by **age**

Holidays	Age 00-49 (95% CI)	Age 50-69 (95% CI)	Age 70+ (95% CI)
New Year's Day	-	-	1.87 (1.58, 2.21)
Good Friday	-	0.82 (0.69, 0.98)	-
US Memorial Day	0.04 (0.01, 0.14)	0.05 (0.02, 0.10)	0.08 (0.04, 0.15)
Independence Day	-	0.03 (0.01, 0.09)	0.08 (0.04, 0.18)
Thanksgiving Day	0.02 (0.00, 0.12)	0.01 (0.00, 0.05)	0.08 (0.04, 0.16)
Black Friday	0.25 (0.15, 0.42)	0.34 (0.26, 0.45)	0.28 (0.19, 0.41)
Christmas Eve	0.22 (0.12, 0.40)	0.25 (0.18, 0.36)	0.15 (0.09, 0.27)
Christmas Day	-	0.04 (0.01, 0.09)	0.10 (0.05, 0.20)

"-" represents dropped nonsignificant holidays.

5. Results (continued)

Table 2: Incidence ratio of holidays to non-holidays by **stage**

Holidays	Localized (95% CI)	Regional (95% CI)	Distant (95% CI)
New Year's Day	0.51 (0.39, 0.67)	0.59 (0.43, 0.86)	-
Good Friday	0.81 (0.69, 0.95)	-	-
US Memorial Day	0.03 (0.01, 0.07)	0.01 (0.00, 0.08)	0.38 (0.17, 0.86)
Independence Day	0.02 (0.00, 0.06)	0.03 (0.01, 0.12)	0.17 (0.04, 0.68)
Thanksgiving Day	0.01 (0.00, 0.04)	0.03 (0.01, 0.09)	0.12 (0.03, 0.48)
Black Friday	-	0.27 (0.18, 0.40)	0.49 (0.24, 0.99)
Christmas Eve	0.1663 (0.11, 0.25)	0.22 (0.13, 0.35)	0.15 (0.04, 0.61)
Christmas Day	-	-	0.16 (0.04, 0.65)

"-" represents dropped nonsignificant holidays.

Table 3: Incidence ratio of holidays to non-holidays by **rural vs. urban**

Holidays	Rural (95% CI)	Urban (95% CI)
New Year's Day	-	-
Good Friday	-	0.83 (0.73, 0.94)
US Memorial Day	0.06 (0.01, 0.44)	0.05 (0.03, 0.09)
Independence Day	-	0.04 (0.02, 0.08)
Thanksgiving Day	0.06 (0.01, 0.45)	0.04 (0.02, 0.07)
Black Friday	0.19 (0.06, 0.58)	0.28 (0.24, 0.34)
Christmas Eve	0.23 (0.07, 0.72)	0.19 (0.14, 0.26)
Christmas Day	-	0.04 (0.02, 0.08)

"-" represents dropped holidays due to nonsignificance
or no incidence to draw statistical conclusions.

Table 4: Seasonality by age, stage and rural vs. urban

Groups	One day	One week	Semi-month	One month	One year
Age 00-49	Yes	No	No	No	Yes
Age 50-69	No	Yes	Yes	Yes	Yes
Age 70+	No	Yes	Yes	Yes	Yes
Localized	Yes	Yes	Yes	Yes	Yes
Regional	Yes	No	No	Yes	No
Distant	No	Yes	No	No	Yes
Unknown/UNS	Yes	No	No	No	No
Rural	No	No	No	No	No
Urban	No	Yes	Yes	Yes	Yes

"Yes" denotes the seasonality in the corresponding column exists while "No" means otherwise;
"Yes" in color orange indicates the strongest seasonality in that group.

6. Discussion

- Timely screening and diagnosis for FBC should be promoted all year round. Public health practitioners can take advantage of major holidays (such as Thanksgiving Day and Memorial Day) and weekends for screening promotion activities.
- Our study reflected whether the seasonality existed and how it differed (relatively strong or weak). For daily time series, resolving the entanglement of multiple seasonalities and trend remains a big challenge. We may turn to structral time series analysis/signal extraction techniques to quantify each seasonal component more appropriately.

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) (NU58DP006299-02/03) 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).