Modeling Breast Cancer Incidence Rates: A Comparison between the Components of Functional Time Series (FTS) Model Applied on Karachi (Pakistan) and US Data

Farah Yasmeen
Department of Statistics, University of Karachi, Karachi, Pakistan
Email: riazfarah@yahoo.com

Received 20 July 2016; accepted 20 August 2016; published 23 August 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

Several studies showed that the breast cancer incidence rates are higher in high-income (developed) countries, due to the link of breast cancer with several risk factors and the presence of systematic screening policies. Some of the authors suggest that lower breast cancer incidence rates in low-income (developing) countries probably reflect international variation in hormonal factors and accessibility to early detection facilities. Recent studies showed that the breast cancer increased rapidly among women in Pakistan (a developing country) and it became the first malignancy among females of Pakistan. Although, the incidence rates may contain important evidence for understanding and control of the disease; however in Pakistan, the breast cancer incidence data have never been available in the last five decades since independence; rather, only hospital-based data are available. In this study, we intend to apply Functional Time Series (FTS) models to the breast cancer incidence rates of United State (developed country), and to see the difference between various components (age and time) of Functional Time Series (FTS) models applied independently on the breast cancer incidence rates of Karachi (Pakistan) and US. Past studies have already suggested that the incidence of US breast cancer cases was expected to increase in the coming decades. A progressive increase in the number of new cases is already predetermined by the high birth rate that occurred during the middle part of the century, and it will lead to nearly a doubling in the number of cases in about 4 decades. We also obtain 15 years predictions of breast cancer incidence rates in United States and compare them with the forecasts of incidence curves for Karachi. Development of methods for cancer incidence trend forecasting can provide a sound and accurate foundation for planning a comprehensive national strategy for optimal partitioning of research resources between the need for development of new treatments and the need for new research directed toward primary preventive measures.

1. Introduction

The carcinoma of breast is a highly heterogeneous disorder, in both etiologically and genetically manners. Approximately 1.38 million new breast cancer cases were diagnosed in 2008 for such commonly occurring cancer worldwide. The breast cancer incidence in Eastern Africa varies from 19.3 new cancer cases per 100,000 women and the incidence rates are very high (as much as 80 new cases per 100,000 women) in the developed countries (except in Japan) [1]. Several studies showed that the incidence rates of breast cancer in most regions of the world were increasing, especially in the developing nations. For example, the incidence rate of breast cancer in Pakistan is about 50 new cancer cases per 100,000 women per year [2] [3].

Most of the diseases of breast take form of a palpable mass, inflammatory lesions or nipple discharge [4]. It is documented that an overall perspective of the various breast problems can be gained by the analysis of a very large series of patients attending outpatient clinics. In the United States, it has been found that about 60% of patients have benign breast disease while only 10% have cancer among them [5]. Breast cancer is the most common malignant tumor in females in Pakistan [6] as well as in US [7]. Breast cancer is the most common cause of death in females in Pakistan while in US, it has now been superseded by lung cancer [8].

Several studies of breast cancer show a higher frequency of this disease in Pakistan when compared to the world standards and to the other studies carried out in Pakistan. The carcinoma of breast is one of the most dreadful diseases of women and the incidence of breast cancer is higher in Pakistani women than other countries of Asia [9] [10]. The KCR (Karachi Cancer Registry) in its report stated that the breast cancer accounted for nearly 34.6% of all cancer cases in the city, making it the most common cancer among the women in Karachi [11].

Breast cancer is most common in developed countries. According to [12], approximately 231,840 new cases of invasive breast cancer were expected in the United States in 2015. About 60,290 new diagnoses of in situ breast cancer were expected among US women in 2015, in addition to invasive breast cancers. The median age at diagnosis for female breast cancer is 61 years [13], and about 12% of women (or 1 in 8) in the US will be diagnosed with breast cancer in their lifetime.

In short, the carcinoma of breast is an enormous public health concern, both in developing and developed countries. Hence, it is necessary to model the breast cancer incidence rates and to see the changes occurring in incidence with the other factors like age and time.

This paper is divided into 6 sections. Section 1 is introductory, while the breast cancer incidence in Karachi and United States are discussed in Sections 2 and 3 respectively. Section 4 is the application of FTS models on breast cancer data of US (White) women while the various components of FTS models for Karachi and US data are compared in Section 5. Finally some concluding remarks are given in Section 6.

2. Breast Cancer Incidence in Karachi

In Karachi, which is the largest and most populated city of Pakistan, the incidence of breast cancer was 69.1 per 100,000 during 1998-2002 [11], with more than half of the breast cancer cases present in advanced stages III and IV [10]. The KCR, Karachi Cancer Registry in its report stated that the breast cancer accounts for nearly 34.6% of all cancer cases in the city, making it the most common cancer among the women in Karachi [11]. The breast cancer incidence rates for the entire city of Karachi have never been calculated; however the relative frequencies of different cancers have been published individually by some radiotherapy centers [14].

[3] applied the FTS (functional time series) models, for the first time to the age-specific breast cancer incidence data of women in Karachi. The secondary data of breast cancer obtained from various locations in Karachi. They include Jinnah Hospital, KIRAN (Karachi Institute of Radiotherapy and Nuclear Medicine) and civil hospital, where the data were available for the years 2004-2011. During this period, a total of 5331 new cases of female breast cancer were registered [14].
The FTS models are first developed by [15], where they applied them to the Australian fertility and French mortality data. These models were also applied to the breast cancer mortality rates (see [16]-[18]). As mentioned earlier, [3] applied for the first time these models to the age-specific incidence rates of breast cancer in Karachi. They used FPC (functional principal components) decomposition [19] to estimate the basis functions \( \{ \phi_i(x) \} \), which are the functions of ages, and the time series coefficients \( \{ \beta_{1i} \} \) described in Equation (3) of [3]. For Karachi, we selected a model with \( k = 4 \) basis functions. The percentages variation explained by these four basis functions were 77.7%, 19.3%, 2.3% and 0.7% respectively. Various components of FTS models for Karachi data are plotted in Figure 1.

For breast cancer incidence data of Karachi, the estimated mean function (see Figure 1) shows that the incidence rates increased with age, and reached their maximum at age 50 years. They decreased slightly after the age of 50 years. The first basis function \( \{ \phi_1(x) \} \), which controls the overall change in trend of breast cancer incidence, represents the younger women (less than 25 years of age).

The plot of first time series coefficient \( \{ \beta_{1i} \} \) corresponds to \( \{ \phi_1(x) \} \) showed fluctuation but overall increase with time (Figure 1). Forecast for this coefficient is clearly showing increasing trend in future. The second basis function explains variation in rates among women aged 30, compared to those who are much

Figure 1. Various components of FTS models applied on breast cancer incidence rates in Karachi. The time series coefficients are plotted, along with 10 years forecasts (2012-2021) of breast cancer incidence, represented by grey shaded region.
younger and much older and the forecast of its coefficient shows a decreasing trend. The third basis function is complex and contrasts those between 20 and 30 or older women (60 years). The third and the higher order bases are usually complex and we do not attempt to interpret them.

The forecasts for the entire incidence curves were also obtained by summing the results after multiplying the forecasts of each coefficient with the basis functions. These forecasts for the breast cancer incidence rates in Karachi were shown in Fig. 4 of [3]. The plot displayed that the future incidence rates will be expected to raise for those aged 50 and above, but the rates will remain stable for those younger than 50 years. The forecast graph suggested that the breast cancer incidence rates among females in Karachi will rise with time in the next ten years (2012-2021), especially for higher age (50 years and over).

3. Breast Cancer Incidence in the United States

Breast cancer is the leading one with 29% of all cancer cases in the United States. Approximately 231,840 new cases of invasive breast cancer and 40,290 breast cancer deaths were expected to occur among US women in 2015 [20]. Several studies have already suggested that the incidence of US breast cancer cases is expected to increase in the coming decades, largely the result of an aging population. A progressive increase in the number of new cases is already predetermined by the high birth rate that occurred during the middle part of the century, and it will lead to nearly a doubling in the number of cases in about 4 decades.

Incidence and Population Data for US

In United States, population-based cancer incidence data have been collected by the National Cancer Institute’s (NCI’s), Surveillance, Epidemiology, and End Results (SEER) Program since 1973. The data have also been collected by the Centers for Disease Control and Prevention’s National Program of Cancer Registries (NPCR) since 1995. The SEER program is the only source for long-term, delay-adjusted, population-based incidence data. Long-term incidence and survival trends were based on data from the 9 oldest SEER areas. They include Connecticut, Hawaii, Iowa, New Mexico, Utah, and the metropolitan areas of Atlanta, Detroit, San Francisco-Oakland, and Seattle-Puget Sound, representing approximately 9% of the US population [21]. From 1992, SEER data have been available for 4 additional populations (Alaska Natives, Los Angeles, San Jose-Monterey, and rural Georgia) that increase the coverage of minority groups. The SEER incidence data for the minority groups cover very small geographical part (only 9%). The Rates/prevalence proportions are available for white, black, other including American Indians and Alaska Natives. The data contains one record for each of 4,863,414 tumors [21]. The data are available for 19 age groups (<1 year, 1 - 4 years, 5 - 9 years, - ... , 85+ years) and single ages with 85+. However in this study, we analyzed the data for fifteen age groups: 15 - 19, 20 - 24, 25 - 29, 30 - 34, 35 - 39, 40 - 44, 45 - 49, 50 - 54, 55 - 59, 60 - 64, 65 - 69, 70 - 74, 75 - 79, 80 - 84 and 85+. We investigated the newly diagnosed cases and incidence rates of breast cancer among these fifteen age-groups, as the incidence rates among the other four groups (<1 year, 1 - 4 years, 5 - 9 years and 10 - 14 years) are extremely small or negligible.

For this study, we obtained the data of breast cancer incidence of US-Whites (female only) from the SEER program, using SEER*Stat software. Here we intend to compare the incidence rates among women in Karachi and US-Whites. In a forthcoming paper, we will discuss the application of FTS model to breast cancer incidence among Black and White women in United States.

For this paper, all statistical analyses were performed in R version 3.3.1 using the R packages demography [22] and forecast [23], available at CRAN.


Figure 2 depicts the observed breast cancer incidence rates, for various age-groups for US-White women during 1973-2013. The rainbow plot [24] is used to represent different age-groups. Using this plot, the incidence curves are plotted in a rainbow order with the earlier age groups (15 - 19) is represented by red and the last group (85+) is plotted by violet color. The other age groups are appeared in the rainbow order (see [24] for details). The R-package “rainbow” is used to construct this plot.

From Figure 2, we can say that the breast cancer incidence rates were increasing with age monotonically, with relatively smaller rates for women under 40 years of age and higher for the women ages 40 years and above. It is important to note that the rates were increasing since 1973 till 1999, but they increased very sharply from...

1999 to 2000 for all age-groups. It may be due to increase use of mammography during this time period. The rates increased very sharply for women ages 55 years and above, from 1999 to 2000, and then they declined for 2002-2003. After that period, they increased slightly for age groups 65 - 69 and 85+, and decreased for all other age groups 55 years and older.

In order to analyse the incidence rates and to describe the changes in incidence rates with age, we also plot the incidence curves (as function of age) for selected years in Figure 3. The years are 1980, 1985, 1990, 1995, 1999, 2000, 2005, 2010 and 2013. We take a 5-year interval, except for 1999 (the year after which the incidence rates increased sharply) and 2013 (the last available year for the data).

The next step is to obtain functional observation by applying smoothing. In Figure 4, original log incidence rates and rates after smoothing are plotted. Penalized regression splines are used to smooth these rates. The graph shows increasing pattern especially for 50 years and above. Then we apply FTS model [15] to these smoothed curves. A model with 4 basis function is used, however, only 3 basis functions are sufficient as the percentage variation explained due to first 3 bases is 97.4%, 2.2% and 0.3% respectively, leaving only 0.01% variation for rest of the components.

5. Comparison of Different Components of FTS Models

After applying FTS model [15] to the US data, we obtain the mean function and the three basis functions \( \{ \phi(x) \} \), along with their time series coefficients \( \{ \beta_i \} \). All these components, with 15-step ahead forecasts of the time series coefficients are plotted in Figure 5.

The mean function for US data shows that the breast cancer incidence rates increased with age monotonically. The first basis function represents those US women who are very young (under 40 years of age and women aged 65 years). The first time series coefficient shows that the incidence rates of those women were increasing since 1973, they increased very sharply from 1999 to 2000, and then they increase relatively slowly from 2000-2013. 15-year forecast of this coefficient shows that future incidence rates are expected to increase for young US White females (under 40 years).

The second basis function for US data (from Figure 5) represents older women (50 years and above) and corresponding time series coefficient shows that the incidence rates among older US women were decreased since 1973 to 1980, then increased from 1980 to 1995 and remain nearly stable during 1995-2005. The breast cancer incidence rates were decreased from 2005 till 2013 and future incidence rates in this age group are expected to decrease relatively slowly in the next 15 years. The other components of FTS model are relatively complex and we do not attempt to interpret them.

Forecasts for the entire incidence curves are obtained by summing the results after multiplying the forecasts

Figure 4. Log incidence rates for US-Whites (females) during (1973-2013) as functional observations. On the left panel, original log incidence rates are plotted whereas the rates after smoothing are plotted on the right panel. Penalized regression splines are used to smooth the log incidence rates.

of each coefficient with the basis functions. These forecasts for the breast cancer incidence rates in US for next 15 years (2014-2028) are shown in Figure 6. This plot displays that the future incidence rates will rise for all age-groups. The expected increase in future breast cancer incidence rates is higher for older ages (55 years and above) and relatively slower in the young age-groups (under 45 years).

6. Conclusions

Several studies showed that the breast cancer incidence rates are higher in developed countries, due to its links with several risk factors [25] and the presence of systematic screening policies [26]. It is also observed a decreasing trend occurred in incidence rates since 2003 due to the decreasing use of Hormone replacement therapy (HRT) [27]. Some authors suggest that lower breast cancer incidence rates in low-income (developing) countries
probably reflect international variation in hormonal factors and problems in accessibility to early detection facilities [28] [29].

In this paper, we applied functional time series (FTS) models to the breast cancer incidence data of women in the United States (Whites only). 15-year forecasts of breast cancer incidence rates among US-White women are also obtained. It is found that in the next 15-year from the last available year 2013 (i.e. during 2014-2028), the future breast cancer incidence rates among US-White females are expected to increase for all age-groups. This rise in future rates is expected relatively slower for young white females (aged 45 years or lower,) and higher for
older age females (55 years and above).

By plotting the incidence data obtained from SEER program, we found that the breast cancer incidence rates among US-White females have been increased very sharply from 1999 to 2000. Although, this sharp increase in incidence rates has not been found in literature, however; the estimated number of new cases of breast cancer among US females were 178,700 in 1998, 175,000 in 1999 and 182,800 in 2000 according to various publications of American Cancer Society (ACS) (see [30]-[32]). It shows an increase of about 7800 new case from year 1999 to year 2000.

Currently, the use of mammography is the most effective way of detecting breast cancer at early stages in case when this disease is most treatable. Over the past decades, mammography screening rates in the United States have remained fairly stable, and screening rates remain relatively low for some of the groups [33]. For targeting the populations at-risk, breast cancer incidence trends along with continued monitoring of breast cancer screening practices are important. Early detection and effective interventions are also play a major role in order to improve breast cancer prevention.

We also compared various components of FTS model for US data, with those obtained from the application of FTS models to the breast cancer incidence data of Karachi (Pakistan). In contrast to the US data, the forecast plot for breast cancer incidence rates in Karachi, obtained in [3], showed that the future incidence rates will be expected to rise only for those aged 50 years and above, but the rates will remain stable for the women younger than 50 years.

To consider age-effect and the effect of other variables, [34] carried out a study at Jinnah Postgraduate Medical Centre, Karachi-Pakistan for a period of three years. The purpose was to evaluate the frequency world standards and to other studies carried out in Pakistan. The patients were between 11 - 90 years of age, however, all cancers were detected between ages 20 - 90 years. A total of 300 patients attending breast care clinics were evaluated and were followed up till biopsy. Majority of the patients belonged to the lower socioeconomic group. Data were collected through a specifically designed Performa. It has been found that most of the females were married (69%) and majority of them (67%) was in the reproductive age, while 33% were postmenopausal. In this study, it was found that breast cancer has a higher frequency of this disease in Pakistan when compared to the world standards. An interesting result is that the majority of females (78%) were house wives, and getting less chance of exercise in their daily life.

It is documented that high mortality due to breast cancer can be reduced by early diagnosis and early treatments [35]-[36]. In literature, the importance of triple diagnosis including clinical examination, mammography and fine needle aspiration cytology [37] and mass population screening with modern mammography has been suggested [38].

Finally, when compared to the world standards, the pattern of breast diseases in Pakistan shows a significant difference. In Pakistan, it has been reported that only 1% of the patients come up having no significant breast disease while in USA, 30% of the patients reporting for checkup have no breast disease [39]. This difference indicates lack of awareness, screening facility and lack of education in Pakistani setup. It has been found that the incidence of fibrocystic disease is much higher in developed countries because of more females getting hormone replacement therapy, as this disease is estrogen dependent. Apart from them, the inflammatory diseases are more common in Pakistan (and other developing countries) due to poor hygiene and low socioeconomic status. Majority of inflammatory diseases are related to lactation and trauma. Breast cancer is the most common disease in Pakistan and other and low-income countries, hence, there is a need of breast care clinics and population screening programs in developing countries to detect breast cancer at an early stage to provide treatment in order to reduce the high mortality from this disease.

References


Nutrition.


