Research Article

Forecasting Age-Specific Brain Cancer Mortality Rates Using Functional Data Analysis Models

Keshav P. Pokhrel¹ and Chris P. Tsokos²

¹Department of Mathematics and Computer Systems, Mercyhurst University, 501 East 38th Street, Erie, PA 16546, USA

²Department of Mathematics and Statistics, University of South Florida, 4202 E Fowler Avenue, Tampa, FL 33620, USA

Correspondence should be addressed to Keshav P. Pokhrel; kpokhrel@mercyhurst.edu

Received 30 July 2014; Revised 2 January 2015; Accepted 20 January 2015

Academic Editor: Peter N. Lee

Copyright © 2015 K. P. Pokhrel and C. P. Tsokos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Incidence and mortality rates are considered as a guideline for planning public health strategies and allocating resources. We apply functional data analysis techniques to model age-specific brain cancer mortality trend and forecast entire age-specific functions using exponential smoothing state-space models. The age-specific mortality curves are decomposed using principal component analysis and fit functional time series model with basis functions. Nonparametric smoothing methods are used to mitigate the existing randomness in the observed data. We use functional time series model on age-specific brain cancer mortality rates and forecast mortality curves with prediction intervals using exponential smoothing state-space model. We also present a disparity of brain cancer mortality rates among the age groups together with the rate of change of mortality rates. The data were obtained from the Surveillance, Epidemiology and End Results (SEER) program of the United States. The brain cancer mortality rates, classified under International Classification Disease code ICD-O-3, were extracted from SEER*Stat software.

1. Introduction

Functional data analysis is about the analysis of information on curves or functions [1]. Functional data is multivariate data with an ordering on the dimensions Muller, 2006. In the present study we are interested in the distribution of functions means, covariances, and relationships of functions to certain responses and other functions. One of the major advantages of functional data analysis is a strong possibility of using rate of change or derivatives of curves. In this study we combine methodologies from functional data analysis, nonparametric statistics, and time series forecasting. There has been an intense research in this area during the last decade [2]. In addition, future mortality rates are of great interests for strategic planner and the insurance industry. Accurate forecasting can be a strong indicator for allocating budget, planning, and policy making. Here, we develop forecasting models for crude mortality rates of malignant primary brain and central nervous system cancer of the United States for the study period of 1969–2008.

The age-specific mortality rates can play a significant role for the allocation of resources for brain tumor control and

evaluations. Primary purpose of this study is to apply functional data analysis techniques to model age-specific brain cancer mortality time trends and forecast entire age-specific mortality function using state-space-approach [3]. Our aim in the present study is to answer the following questions:

- (i) to capture the subtle pattern of variation in mortality,
- (ii) to find the prediction interval of the forecast,
- (iii) to find a flexible model so that it could incorporate the covariates such as screening and treatment effects into the analysis.

However, there are some limitation in functional time series forecasting model. It only incorporates year of death (calendar period) and fails to consider year of birth (cohort) effects. The calendar period incidence and mortality trend can reveal the effects of new medical interventions but fail to reflect changes in risk factors such as screening and radiation therapy.

Currently, most of the forecasting models are based on age-period-cohort methods. These methods are structured to estimate the mortality rates of breast cancer [4], prostate

cancer [5], and cervical cancer [6, 7]. These methods are basically regression models with mortality or incidence rates as outcome variables using Poisson error distribution with log link function. The most common problem in these models is nonidentifiability of parameters, very strong parametric assumptions, and sensitivity of projections and lack of inclusion of most recent changes in cohort effects. Most importantly, a limitation of these studies is that none of them forecast mortality rates with age-related changes.

The linear extrapolation and nonlinear Poisson distribution models are discussed in [8, 9] whereas a Bayesian age-period-cohort model with autoregressive smoothing of each of age, period, and cohort components is studied such that the resulting projections are estimated from current and past trends of the data in [4]. Lee and Carter (LC) [10] method is one of the most influential methods in demographic forecasting. This is perhaps the most cited paper by demographic researchers. LC proposed a long term forecasting method to extrapolate mortality rates and applied it to forecast US mortality rates for the year 2065.

There has been numerous extensions of LC method, some of the extensions and modifications of LC method can be found in [11]; Renshaw and Haberman (2003) [12], and the applications of LC method in fertility forecasting can found in Lee [13]. The method proposed by Lee and Carter in 1992 has become the leading statistical model of mortality/forecasting in the demographic literature [14], Deaton and Paxson, 2004 [15]. It was used as a benchmark for recent Census Bureau population forecasts [16], and two US social security technical advisory panels recommended its use, or the use of a method consistent with it [11].

A comprehensive discussion of the patterns of mortality rates for then G 7 countries is presented by Tuljapurkar et al. [14] using LC method. The LC model predicted 1-to-4-year higher life expectancy than official projections in the industrial nations, with larger differences for Japan.

There are numerous uncertainties which affect the mortality rates; however a probabilistically sound forecasting method, like LC method, is particularly useful to address the long term funding problems of public pension and insurance for increasingly ageing population in the industrial world.

2. Functional Data Analysis (FDA) Model

2.1. An Overview. The Lee Carter model for age-specific mortality rates is given by

$$\ln(m_{x,t}) = a_x + b_x k_t + \epsilon_{x,t}, \quad (1)$$

where a_x is general age shape of age-specific mortality rates, b_x represents the tendency of mortality at age x , and k_t is the time varying index. Equation (1) is a linear model of an unobserved period-specific intensity index, with parameter depending on age (LC 1992). LC model uses singular value decomposition (SVD) method for exact least square fit; however, a simple linear regression method can also approximate the parameters. LC incorporates a random walk with drift for the time series formed by k_t , which is expressed as

$$k_t = k_{t-1} + c + e_t, \quad (2)$$

where c is the drift term, k is forecast to decline linearly with increment of c , and e_t are permanently incorporated in the trajectory [11]. The standard error of c could be used for the detailed measure of uncertainty in forecasting k .

Generalization. The Hyndman-Ullah [3] approach is a generalization of Lee and Carter (LC 1992) method.

Our primary goal is to find functional forecasting model for mortality rates of brain and central nervous system tumor in the United States. The proposed forecasting model is developed in the realm of functional data [1] for modeling log mortality rates. To develop the functional data, we invoke the nonparametric smoothing methods to mitigate the existing randomness in the observed information. In addition, the problems related to age groups and issues of outlying years are reasonably addressed by using functional principal component [1, 3]. The observed data is smoothed and principal component analysis is applied after smoothing the observed data.

The forecasting methodology by Hyndman-Ullah is a generalization of Lee and Carter method. This approach uses functional data analysis techniques and treats the age-specific mortality curves as the units of analysis rather than the discrete observations [3]. In practice, functional data are usually observed and recorded discretely as n pairs (t_j, y_j) , and y_j is a snapshot of the function at time t_j , possibly blurred by measurement error [1]. Generalized Lee Carter method models the mortality rates as a continuous function of age and captures the subtle variation between years. In addition the smoothness of the data reduces the observational error and forecast the entire function with prediction intervals [3].

In the following section we discuss the use of more flexible [3] method to model the brain cancer mortality rates which uses multiple functions to capture the changes in rates.

2.2. FDA Model for Mortality Data. Let $m_t(x)$ denote the mortality rate for midpoint of age group x and year t , $t = 1, \dots, n$. We model the log mortality,

$$y_t(t) = \log[m_t(x)], \quad (3)$$

and assume that there are underlying functions $f_t(x)$ that we are observing with error [17]. The mortality rates as a smooth function of age can be expressed as

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i) \epsilon_{t,i}, \quad (4)$$

where x_i is the center of age group i ($i = 1, \dots, p$), $\epsilon_{t,i}$ is an independent and identically distributed standard normal random variable, and $\sigma_t(x_i)$ allows the amount of noise to vary with the age x . After developing functions of the given mortality rates, we fit the model

$$f_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x), \quad (5)$$

where $\mu(x)$ is the mean log mortality rate across years, $\phi_k(x)$ is a set of orthogonal basis functions, and $e_t(x)$ is the model error which is assumed to be serially uncorrelated [17].

The mean log mortality rates $\mu(x)$ are estimated by using penalized regression splines (Wood, 2000) [18]. The pairs $(\beta_{t,k}, \phi_k(x))$ for $k = 1, 2, \dots, K$ are estimated by decomposing the data into principal components, whereas $e_t(x)$ is the difference between spline curve and fitted curve from the model. We wish to estimate the optimal set of K orthogonal basis functions. The optimal orthogonal basis function $\phi_k(x)$ is obtained via principal components (see Ramsay and Silverman, 2005, pages 151-152). Specifically, for a given K , we want to find the basis functions $\{\phi_k(x)\}$ which minimizes the mean integrated squared error:

$$\text{MISE} = \frac{1}{n} \sum_{t=1}^n \int e_t^2(x) dx. \quad (6)$$

This is achieved using functional principal components (FPC) decomposition, [19], applied to the curves $\{f_t(x)\}$ which provides the least number of basis functions, and explores the coefficients which are uncorrelated with each other.

2.3. *Forecasting.* Equations (4) and (5) together yield

$$\begin{aligned} y_t(x_i) &= \ln[m_{x,t}] \\ &= \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x) + \sigma_t(x_i) \varepsilon_{t,i}. \end{aligned} \quad (7)$$

Let $\hat{\beta}(n, k, h)$ denote the h -step forecast of $\beta_{n+h,k}$ and let $\hat{f}_{n,h}(x)$ denote the h -step ahead forecast of $f_{n+h}(x)$. Then,

$$\hat{f}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}(n, k, h) \hat{\phi}_k(x). \quad (8)$$

To forecast the coefficients from (8) we use state-space model for exponential smoothing. The exponential smoothing method provides a statistical framework for automatic forecasting [19]. This forecast then multiplied with estimated basis function to obtain the forecast of the entire function. In addition, exponential smoothing techniques also provide prediction intervals for the forecast by incorporating variance of error terms [20]. Forecast from exponential smoothing methods is estimated recursively where recent observations are given more weight than historical data. This method accommodates additive and multiplicative trend with automatic model selection for the given time series [17].

The state-space models provide a convenient and powerful framework for analyzing sequential data (see Harvey 1989) [21]. Many mortality data sets require extrapolation, as data has a time dimension. The state-space model can be used to calculate smooth feature or signals and associated standard errors provided the model is of the state-space form.

The sum of the variances of all individual terms is the forecast variance [3]:

$$\begin{aligned} \text{Var}[\hat{y}_{n+h,k}(x)] \\ = \hat{\sigma}_\mu^2(x) + \sum_{k=1}^K u_{n+h,k} \hat{\phi}_k^2(x) + v(x) + \sigma_t^2(x), \end{aligned} \quad (9)$$

where $\hat{\sigma}_\mu^2(x)$ is the variance obtained using the smoothing method. The forecast variance is given by

$$u_{n+h,k} = \text{Var}(\beta_{n+h,k} | \beta_{1,k}, \dots, \beta_{n,k}), \quad (10)$$

$v_k(x)$: sum of square of residuals, $\sigma_\mu^2(x)$: variance of the smooth estimate $\hat{\mu}(x)$, $\sigma^2(x)$ is estimated by assuming binomial distribution of $m_t(x)$, and $v(x)$ is the mean of $\hat{e}_t^2(x)$ for each x .

We evaluate the accuracy of the mortality forecast by computing the mean integrated squared forecasting error (MISFE) defined as

$$\text{MISFE}(h) = \frac{1}{n-m+1} \sum_{t=m}^n \int [y_{t+h}(x) - \hat{f}_{t,h}(x)]^2 dx, \quad (11)$$

where m is the minimum number of observations used in fitting the model.

3. Data

An estimated 69,720 (10% increment from 2010) new cases of primary nonmalignant and malignant brain and central nervous system tumors are expected to be diagnosed in the United States in 2013 [22]. This caused 13,700 (5.5% increment from 2009) deaths because of the primary malignant brain and central nervous system tumors in the United States in 2012. It is estimated that 24,620 men and women (13,630 men and 10,990 women) will be diagnosed and 13,140 (4.26% increment from 2010) men and women are estimated to be deceased of brain and other nervous system cancer in 2013. Males and females have a 0.7% and 0.6% lifetime risk of being diagnosed with a primary malignant brain/central nervous system tumor. These projections are of major public health interest. However, their interpretation may be complex because of the effect of screening, risk factors, and accessibility of effective treatments.

Crude mortality rates per 100,000 persons based on the 2000 standard US population were extracted using the SEER*Stat 7.0.5 software of the Surveillance Epidemiology and End Results program, National Cancer Institute Institutes. We are using 416,480 (229,467 males and 187,013 females) malignant Brain cancer patients, where 381,238 are whites, 24,336 are African Americans, and 4,891 are others (American Indian/AK Native, Asian/Pacific Islander, 1969–2008).

The mortality rates were at their highest from 1885 to 1995. After 2000 we observed that the overall rates are leveling off or declining. From 2003–2008, the median age at death for cancer of the brain and other nervous system was 64 years of age. Approximately 4.2% died under age of 20; 3.8% died between 20 and 34; 7.1% between 35 and 44; 14.9% between 45 and 54; 21.8% between 55 and 64; 22.2% between 65 and 74; 19.6% between 75 and 84; 6.3% died at 85+ years of age. For the first part of our study, we use annual crude mortality rates in the United states from 1969 to 2008 in 5-year age groups (01–04, 05–09, ..., 80–84, 85+).

FIGURE 1: Age-specific brain cancer mortality by age group among total males (1969–2008). In the figure 49 represents the age group 45–49, 64 represents the age group 60–64, and 89 represent the age group above 85.

FIGURE 2: Age-specific brain cancer mortality by age group among total females (1969–2008). In the figure 49 represents the age group 45–49, 64 represents the age group 60–64, and 89 represent the age group above 85.

4. Results

For this study we obtained the data from the Surveillance, Epidemiology and End Results (SEER) program of National Cancer Institute in the United States [23]. Specifically, mortality data were obtained from the National Center for Health Statistics (NCHS) available on the SEER*Stat database. Annual age-specific brain cancer mortality data are designated by ICD 8 & 9 (1979–1998) code 174 and ICD 10 (1999+) codes C70 and C71. The data available is about different racial subgroups since 1969 in nineteen year age groups: 01–04, 05–09, 10–14, 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+.

Figure 1 displays brain cancer mortality rates in US male by age group for the period of 1969–2007. We observe that the mortality rates for ages below 40 years show no obvious trend; for the same period the pattern of mortality rates for ages more than 40 exhibited significant variation of mortality rates for the elderly. The graph shows the brain cancer mortality rates among males between 45 and 65 years declined slowly throughout the study period. The pattern for the elderly (aged 65 and above) population is clearly increasing from 1969 to 2000 and started declining after 2000. However, mortality rates for the population subgroups between 45–65 show slightly decreasing pattern. The mortality rates for the age groups 80–84 and 85+ are very unstable and difficult to interpret because of the limited availability of data.

In Figure 2, we present the age-specific mortality rates with respect to the year of decease for the female subpopulation. The trends in mortality rates do not show any visible pattern for the age groups less than 40 years. The rates for the

age groups 50–54, 55–59, and 60–64 show decrease in mortality during the whole period of study. Overall, the mortality rate for males is lower than that of female but the age-specific mortality rates show no obvious difference between the rates. Figure 3 displays the age-specific mortality rates with respect to age for the women subpopulation for brain cancer mortality rates during the entire period of study, 1969–2008. The graph depicts similar variation of hazard rates as for the male subpopulation. These rates are notably unstable for the young (less than 20 years), monotonically increasing for middle age population (between 20 and 67), and decreasing with strong instability for the elderly (67 and above). In general, mortality rates are directly proportional to age, and for every year, highest mortality rates are observed in the age groups 80–84, 85 and above; this is true for both male and female subpopulation.

In Figure 4, we present the mortality rates of the male subpopulation for the entire period of study. The male brain cancer mortality rates are higher than female in magnitude but show similar pattern for different clusters of the population. Figure 4 clearly indicates three different patterns of mortality rates: first, unstable and decreasing mortality rates for patients below 20; second, reasonably stable pattern of mortality rates for middle aged population subgroups (ages 20–65); and third, significantly increasing and very erratic behavior of mortality rates for the elderly (ages 65 and above). Three different patterns of mortality rates by age are present for the total population different subgroups of population as well. However, this behavior of mortality rates is more apparent in male population.

FIGURE 3: Age-specific brain cancer mortality with respect to age for females subgroup of the population (1969–2008).

FIGURE 4: Brain cancer mortality rates as a function of age for male subpopulation (1969–2008).

In the next section, we develop smooth functions of mortality rates as a function of age for different population subgroups. The smooth (or interpolated) functions are developed for brain and CNS tumor mortality rates; these smooth functions are used to find principal components to measure the variability of the data. Penalized regression splines are used to find smooth functional curve with a monotonicity constraint. The curves are monotonically increasing for age greater than 20. The smoothness of the curves is controlled by a smoothing parameter λ based on Wood [18]. The term functional refers to the intrinsic structure of the data under the assumption that there exists a function of x giving rise to the observed data [1].

For the given period of study 1969–2008, we assume the existence of the function of age with respect to the mortality rates plus observational error. The observation errors can be mistakes in collecting or recording the data. The observed data are smoothed using penalized regression spline with penalty $\lambda = 4$. We assume that $f_t(x)$ is monotonic for $x > 20$ years. It is reasonable to assume monotonicity in mortality data (the older you are, the more likely you are to die). The data are technically smooth after 20 years and there are no outlying years present. We also analysed the same data using linear interpolation as well; final results were not significantly different from the results obtained from regression splines. We would like to develop a mathematically independent set of basis functions ϕ_k which can be approximated by linear combination of K of these functions. The basis system is not specifically defined by the fixed number of parameters but rather itself is a parameter that we choose as per the nature of the data. One of the primary criteria to choose basis is to test the behavior of their derivatives; some of the bases that give reasonable fit may end up with poor derivative estimates. We choose a model with $K = 6$ basis functions. The number of basis functions is chosen by minimizing the mean square

error (MSE) and mean integrated squared forecasting error (MISFE) [17, 19]. Six basis functions ended up with minimum mean square error (MSE) together with mean integrated squared error and mean integrated squared forecasting error (MISFE). No additional basis function contributed in further reduction of MISFE. The estimated coefficients are presented in the bottom of Figure 5.

The functional regression model with six basis functions explained 98.8% of the variability of the data for total observation (both male and female) during the period of study in the United States. The proportion of variation explained by each basis function in the decreasing order of their magnitude is 92.0%, 4.8%, 0.6%, 0.6%, 0.6%, 0.3%, and 0.5% for $k = 1, \dots, 6$, respectively. The mean square error (MSE) for the variability 0.00082 and the integrated squared error (ISE) is given by 0.07217. For females, the first six basis functions in the functional model explain 87.6%, 5.5%, 1.8%, 1.1%, 0.8%, and 0.7% of the total variation of female mortality rates with mean squared error 0.00183 and integrated squared error 0.16050, while 88.6%, 4.8%, 1.7%, 1.2%, 0.7%, and 0.6% of the total variability of mortality rates are explained for male population with $\text{MSE} = 0.00153$ and $\text{ISE} = 0.13532$. The overall goodness of fit was assessed by using residuals of fitted mortality model using image plots. The plots showed no lack of fit. We also checked autocorrelation in the observational error, $\epsilon_t(x_i)$, for each x_i and in the one step forecast error for various age groups. The autocorrelation was either insignificant or sufficiently small. In addition, we evaluated the accuracy of mortality forecasts by using MISFE where we set $m = 10$ as minimum number of observations used in fitting the model. That is, we fit the model up to time t and predict next n periods to obtain MISFE.

The optimal orthogonal basis functions are computed from principal component decomposition. We want to find a

FIGURE 5: The functional model for brain cancer mortality in the USA. Top line: the mean function and first six basis functions. Bottom line: the coefficients associated with each of the basis functions.

set of exactly K orthogonal functions ϕ_k so that the expansion of each curve in terms of these basis functions approximates the curve as closely as possible (see Ramsay and Silverman, 2005, pages 151-152). Figure 5 explores the first six basis functions together with the corresponding coefficients. The basis functions and the time series coefficients model the overall variability of the mortality rates. The first basis function (ϕ_1) models the higher age groups (around 80 years), as the score is the largest in negative direction; second basis function (ϕ_2) models the middle age (20–65); the third and fourth (ϕ_3 and ϕ_4) basis functions represent the infants and people under 20 years of age respectively. The fifth and sixth (ϕ_5 and ϕ_6) are relatively complex to explain and we are not attempting to explain these functions because of their unpredictable variability.

The plots of time series coefficients, Figure 5, depict a continuously increasing pattern before 1990, while the rates show a declining pattern during the decade of 1990–2000. The first time series coefficients represent decreasing trend but since the basis function has negative sign, the older ages (around 80 years) have been increasing during the study period and it can also be observed from Figure 3. The second time series is first increasing till 1990 and then decreasing which corresponds to the age group 20–40 years. More specifically, the mortality rates are increasing for the age groups more than 65, slightly decreasing for the population with the age between 40 and 65, and the rates are leveled off for younger population. The variability of mortality rates for the patients more than 80 years is remarkably high, which is numerically important and less explained by the basis functions. The erratic death rates for the elderly may be because of the measurement error than due to behavior of the rates. Similar study by Coale and Kisker [24] showed that the mortality data are highly susceptible and fraught with various types of measurement problems. It is more reasonable to have detail and separate study of patients above 80 years; we excluded the mortality rates for the age groups 80–84 and 85+ in the later part of our study.

The forecast of the brain cancer mortality rates is calculated by multiplying the time series coefficients with the basis functions. Figure 6 shows 10-year forecast of mortality rates for male and female together during the period of study. We observe that the brain and CNS cancer mortality rates are expected to increase with respect to age. One-year and ten-year forecast show a declining pattern of mortality rates for brain and CNS tumor patients of all the age groups. A declining pattern of mortality rates is observed for the ages less than 60. However, the declining pattern of mortality rates is inverted for the elderly.

The difference between the mortality rates is remarkably higher in the age groups more than 75 years. Average difference of mortality rates in the age groups 75–79, 80–84, 85–89, and 89+ is 0.05 per 100,000 per year. We observe that the average mortality rates of a person of 62 years of age is almost 10 times higher than that of a person of 32 years of age. Long term forecast shows that the rates are predicted to decline relatively slowly in the next decade. The mortality rates for the total US population are expected to decrease by 1.58% for the age group of 0–4 years, and at the same time

FIGURE 6: Forecast of age-specific mortality of total brain and central nervous system cancer patients in the United States for 2009 and 2018, with 80% confidence intervals.

rates are expected to increase by 5.5% for the age group of 80–84. Specifically, the mortality rates are predicted to increase for all the persons above 65 years of age. For total US population with age less than or equal to 65 the rates are predicted to decrease linearly at the rate of 0.0145 per 100,000 per year (P value < 0.05). We also observed that 20th percentile of difference in mortality rates between 2009 and 2018 is 0.501 per 100,000 per year; 50th percentile is 0.796; and 99 percentile of the difference between mortality rates is 1.57 per 100,000 per year.

In Figure 7, we present one-year and ten-year prediction intervals for male and female populations. The mortality rates are expected to increase, (comparing with 2009) for both the gender by 0.17 (0.37%) persons per 100,000 by 2018. For the same period the mortality rates for males and females separately are expected to increase by 0.33 (0.78%) and 0.11 (0.19%) persons per 100,000 by 2018. This may be because of erratically higher mortality rates of elderly population. However, age-specific forecast for 2009 and 2018 shows slower rate of decline in female mortality rates in comparison to the male population. The average increment in mortality rates is 0.33 persons per 100,000 and 0.84 persons per 100,000 for males and females aged below 65 years, respectively. We also observed that the mortality rate for the age groups more than 65 is increasing in higher rate than other age groups. The average rate of increase in mortality rates for the elderly is 1.23 (5%) and 0.44 (0.38%) persons per 100,000 for males and females, respectively (see Table 1). The mortality rates for the elderly population are subject to error because of availability of information and erratic behavior in the mortality rates. Lee and Carter (1991) [10] also mentioned the unreliability of age groups 85+. In contrast, after clustering the ages into three groups, 0–19 (young adults), 20–64 (middle age), and 65 and above (elderly), we observe that the mortality rates for younger population are estimated to decrease with the highest declining rate followed by the middle aged population.

In Table 1, we present one-year and ten-year forecast of the mortality rates. The average decrease (from 2009 to 2018) in mortality rates for male and female subpopulation is 1.6 and 1.41 persons per 100,000. The variability of the mortality rates for different age groups is clearly noticeable. Despite the

FIGURE 7: Forecast of age-specific mortality of brain and central nervous system cancer in the United States for 2009 and 2018 for male and female subpopulation with 80% confidence intervals.

TABLE 1: Forecast and 80% prediction intervals of mortality rates of total US patients of brain and CNS tumors for 2009 and 2018.

Age	Total						Male						Female					
	2009		2018		2009		2018		2009		2018		2009		2018		2018	
	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI	Mean	80% PI
2	0.62	0.55	0.70	0.61	0.53	0.70	0.66	0.56	0.79	0.65	0.54	0.78	0.57	0.49	0.66	0.56	0.47	0.67
7	0.88	0.80	0.97	0.87	0.78	0.97	0.87	0.76	0.99	0.85	0.74	0.98	0.88	0.78	1.00	0.88	0.77	1.00
12	0.63	0.56	0.71	0.62	0.55	0.71	0.63	0.55	0.74	0.62	0.53	0.74	0.62	0.53	0.71	0.61	0.52	0.71
17	0.48	0.44	0.53	0.48	0.42	0.54	0.53	0.46	0.61	0.53	0.45	0.62	0.42	0.37	0.49	0.42	0.35	0.49
22	0.53	0.48	0.58	0.52	0.47	0.59	0.62	0.53	0.72	0.60	0.51	0.71	0.42	0.34	0.52	0.42	0.33	0.53
27	0.73	0.66	0.80	0.72	0.64	0.81	0.86	0.75	0.99	0.86	0.73	1.00	0.53	0.45	0.63	0.54	0.43	0.66
32	1.10	1.01	1.20	1.09	0.97	1.22	1.33	1.20	1.47	1.32	1.16	1.50	0.84	0.73	0.98	0.84	0.70	1.00
37	1.63	1.51	1.76	1.62	1.46	1.79	2.00	1.83	2.19	1.99	1.78	2.22	1.24	1.11	1.39	1.23	1.07	1.42
42	2.49	2.33	2.66	2.47	2.27	2.70	3.10	2.84	3.39	3.08	2.78	3.42	1.90	1.71	2.10	1.88	1.66	2.12
47	3.70	3.45	3.97	3.67	3.36	4.01	4.58	4.20	4.99	4.54	4.10	5.03	2.87	2.61	3.15	2.85	2.54	3.18
52	5.48	5.17	5.81	5.45	5.02	5.91	6.73	6.25	7.25	6.69	6.10	7.33	4.29	3.94	4.66	4.26	3.82	4.75
57	8.00	7.62	8.41	7.95	7.40	8.54	9.73	9.13	10.38	9.69	8.93	10.50	6.29	5.84	6.78	6.26	5.67	6.91
62	10.84	10.30	11.41	10.80	10.05	11.60	13.25	12.50	14.06	13.27	12.33	14.27	8.56	7.88	9.30	8.55	7.68	9.53
67	13.88	13.20	14.61	13.91	12.91	14.99	16.83	15.71	18.02	17.01	15.59	18.56	11.13	10.32	12.00	11.17	10.10	12.37
72	17.14	16.05	18.31	17.40	15.70	19.28	20.64	19.15	22.26	21.21	19.11	23.54	14.18	12.93	15.54	14.37	12.62	16.36
77	19.22	17.62	20.96	19.77	17.10	22.86	23.81	21.65	26.19	24.87	21.76	28.42	15.59	13.90	17.48	15.98	13.31	19.19
82	19.94	17.30	22.98	21.05	16.60	26.69	24.69	20.94	29.11	26.64	21.17	33.51	16.66	14.10	19.69	17.35	13.12	22.95
87	17.44	14.63	20.78	18.70	13.89	25.19	23.30	19.00	28.57	25.72	19.24	34.38	14.52	11.82	17.82	15.38	10.89	21.73

fact that brain and CNS cancer is one of the most vulnerable cancer for the younger population, the mortality rates are declining faster than middle aged and elderly population. In a separate study of mortality rates for the age groups 0–19 and 20–64 we report smaller MSE as well ISE in comparison to the models with age groups together (see Appendices A and B for forecast of the mortality rates for age groups 0–19 and 20–65).

In Tables 2 and 3, we present predicted change in percentage of mortality rates between 2009 and 2018. Table 2 shows the mortality rates are increasing for the elderly age group whereas the case is inverted for other two age groups. In addition, we can expect that 50% or more younger patients will see 5.14% reduction of mortality rates in the younger subpopulation in 2018 whereas 50% of the middle aged patient (20–65) are predicted to have 3.35% reduction in mortality

TABLE 2: Summary of differences (in %) in mortality rates between 2009 and 2018 for all patients. Differences are noted with – and + indicating decreasing and increasing pattern of mortality.

	Before clustering			After clustering	
	0–19	20–64	65+	0–19	20–64
Median	–1.22	–0.63	+2.89	–5.14	3.35
Maximum	–1.57	–0.86	+7.26	–6.03	–7.77
Minimum	–0.94	–0.11	+0.11	–4.7	–0.459

TABLE 3: Summary of differences (in %) in mortality rates between 2009 and 2018 for male and female subpopulation. Differences are noted with – and + indicating decreasing and increasing pattern of mortality.

	Male		Female	
	0–19	20–64	0–19	20–64
Median	–5.5	–2.5	–4.8	–4.6
Maximum	–6.18	–9.76	–5.8	–6.01
Minimum	–5.1	+0.29	–2.5	–1.7

rates. These percentage changes will be lower if we include the elderly subpopulation (65 and above) in the study.

In Table 3, we present differences in percentage of mortality rates after clustering by sex. The mortality rates are expected to decrease for younger and middle age subpopulation. Interestingly, the rates are predicted to decrease by higher percentage for younger male subpopulation whereas the female middle aged subpopulation are predicted to have higher decrease than their male counterparts.

5. Discussion

In this study of brain cancer mortality in the United states, we present an application of new forecasting method [3] in brain cancer mortality rates. The mortality-age relationship is modeled by using basis expansion of the data which highlights important trends during the period of study. We report that the mortality rates are predicted to decrease continuously for all the brain and CNS patients of age groups 0–19 and 20–64 (predicted to decrease by 3.5% in 2018). Higher mortality rates are estimated for the elderly (65 and above) but these rates are heavily influenced by low sample size and higher fluctuation of mortality rates of the subject age group. We also observed that male population possess relatively homogenous pattern of mortality rates in comparison to female subpopulation. In addition, males have consistently higher mortality rates than female mortality rates.

Our model predicts a decreasing pattern of mortality in brain cancer in the United States for the next decade. We also report relatively small change in mortality for the age group 20–65, while the rates are expected to increase for patients more than 65 years of age. Interestingly, we observed a higher decline in mortality rates for the younger population subgroup. Advancements in treatment have significantly helped to increase survival rates for children with brain tumors [25].

There are three age groups with distinct mortality patterns: young adults (0–19); the middle aged (20–64); and the elderly (65 and above). Predicted rates are reported to be decreased by higher percentage after studying the clusters separately. Before clustering, for young adults, the median of percentage difference between mortality rates is predicted to be lowered by 1.22% in 2018. Similarly, after clustering, the rates are predicted to decrease by 5.14% in 2018. It is also reported that about 75% of children survive at least 5 years after being diagnosed with a brain tumor [26]. Nevertheless, many childhood brain tumor survivors are at risk for long-term neurological complications.

The functional form of the data is able to capture the subtle variations in mortality rates; thus the models and forecasts both have notable strength demographic forecasting [17]. The other studies in modeling mortality and fertility forecasting also have acknowledged the implication of functional data analysis approach in modeling all causes of mortality rates [11, 27]. Functional data analysis method on fertility and mortality data achieves better forecasting results than other approach to mortality forecasting [3]. Moreover, FDA technique allows the variation of mortality rates so that for different ages mortality declines or increases at different rates.

FDA techniques are getting popularity since last decade, especially after the publication of a pioneering book in FDA by Ramsay and Silverman in 2005. This technique has number of strengths in modeling high dimensional and missing data. For a detailed review of FDA and its application in the different fields of study see [2]. Functional analysis techniques make no parametric assumptions about age or period of effects; variation of mortality rates is presented with respect to time so that for different ages the fluctuation of rates can be modeled. Also, functional models are free from the strong parametric assumptions of error variations and do not assume linear dependency between the variables age, period, and cohort. To the best of our knowledge, no other studies modeled or forecast the brain cancer mortality rates with age as functional covariates over time.

Appendices

A. Appendix A

See Table 4.

B. Appendix B

See Table 5.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

TABLE 4: Age- and gender-specific prediction of brain cancer mortality rates of young adults for the years 2009 and 2018. In the first table, prediction is made for all the patients under study. In the second table, prediction of mortality rates is made separately for young adults (0–19 years). Prediction of mortality rates show higher decreasing rate if analysis is done after clustering the data by age groups.

(a)

Age	Total						Male						Female					
	2009			2018			2009			2018			2009			2018		
	Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI	
2	0.62	0.55	0.70	0.61	0.53	0.70	0.66	0.56	0.79	0.65	0.54	0.78	0.57	0.49	0.66	0.56	0.47	0.67
7	0.88	0.80	0.97	0.87	0.78	0.97	0.87	0.76	0.99	0.85	0.74	0.98	0.88	0.78	1.00	0.88	0.77	1.00
12	0.63	0.56	0.71	0.62	0.55	0.71	0.63	0.55	0.74	0.62	0.53	0.74	0.62	0.53	0.71	0.61	0.52	0.71
17	0.48	0.44	0.53	0.48	0.42	0.54	0.53	0.46	0.61	0.53	0.45	0.62	0.42	0.37	0.49	0.42	0.35	0.49

(b)

Age	Total						Male						Female					
	2009			2018			2009			2018			2009			2018		
	Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI	
2	0.63	0.56	0.71	0.59	0.52	0.67	0.66	0.56	0.79	0.62	0.52	0.74	0.59	0.51	0.69	0.56	0.48	0.65
7	0.84	0.76	0.92	0.80	0.73	0.89	0.81	0.71	0.93	0.77	0.67	0.88	0.86	0.76	0.97	0.84	0.74	0.94
12	0.61	0.55	0.68	0.58	0.52	0.65	0.64	0.55	0.74	0.60	0.52	0.70	0.56	0.49	0.65	0.54	0.47	0.62
17	0.50	0.45	0.55	0.47	0.42	0.52	0.57	0.49	0.66	0.54	0.46	0.62	0.42	0.36	0.48	0.39	0.34	0.46

TABLE 5: Age- and gender-specific prediction of brain cancer mortality rates of young adults for the years 2009 and 2018. In the first table, prediction is made for all the patients under study. In the second table, prediction of mortality rates is made separately for middle age (20–64 years) subpopulation. Prediction of mortality rates show higher decreasing rate if analysis is done after clustering the data by age groups.

(a)

Age	Total						Male						Female					
	2009			2018			2009			2018			2009			2018		
	Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI	
22	0.53	0.48	0.58	0.52	0.47	0.59	0.62	0.53	0.72	0.60	0.51	0.71	0.42	0.34	0.52	0.42	0.33	0.53
27	0.73	0.66	0.80	0.72	0.64	0.81	0.86	0.75	0.99	0.86	0.73	1.00	0.53	0.45	0.63	0.54	0.43	0.66
32	1.10	1.01	1.20	1.09	0.97	1.22	1.33	1.20	1.47	1.32	1.16	1.50	0.84	0.73	0.98	0.84	0.70	1.00
37	1.63	1.51	1.76	1.62	1.46	1.79	2.00	1.83	2.19	1.99	1.78	2.22	1.24	1.11	1.39	1.23	1.07	1.42
42	2.49	2.33	2.66	2.47	2.27	2.70	3.10	2.84	3.39	3.08	2.78	3.42	1.90	1.71	2.10	1.88	1.66	2.12
47	3.70	3.45	3.97	3.67	3.36	4.01	4.58	4.20	4.99	4.54	4.10	5.03	2.87	2.61	3.15	2.85	2.54	3.18
52	5.48	5.17	5.81	5.45	5.02	5.91	6.73	6.25	7.25	6.69	6.10	7.33	4.29	3.94	4.66	4.26	3.82	4.75
57	8.00	7.62	8.41	7.95	7.40	8.54	9.73	9.13	10.38	9.69	8.93	10.50	6.29	5.84	6.78	6.26	5.67	6.91
62	10.84	10.30	11.41	10.80	10.05	11.60	13.25	12.50	14.06	13.27	12.33	14.27	8.56	7.88	9.30	8.55	7.68	9.53

(b)

Age	Total						Male						Female					
	2009			2018			2009			2018			2009			2018		
	Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI		Mean	80% PI	
22	0.50	0.45	0.55	0.46	0.40	0.52	0.57	0.50	0.67	0.52	0.44	0.61	0.39	0.32	0.48	0.36	0.29	0.45
27	0.71	0.65	0.78	0.68	0.62	0.76	0.83	0.73	0.95	0.82	0.71	0.95	0.55	0.47	0.66	0.52	0.44	0.62
32	1.08	0.99	1.18	1.04	0.94	1.15	1.41	1.28	1.56	1.32	1.18	1.47	0.82	0.71	0.94	0.77	0.66	0.90
37	1.57	1.46	1.69	1.52	1.39	1.66	2.02	1.85	2.20	1.90	1.72	2.10	1.21	1.08	1.35	1.15	1.02	1.30
42	2.41	2.25	2.58	2.33	2.15	2.52	3.03	2.78	3.31	2.99	2.72	3.29	1.81	1.62	2.02	1.73	1.53	1.95
47	3.56	3.32	3.82	3.44	3.17	3.73	4.50	4.14	4.90	4.38	3.99	4.81	2.77	2.51	3.06	2.65	2.37	2.95
52	5.36	5.06	5.67	5.19	4.84	5.57	6.59	6.12	7.09	6.47	5.95	7.03	4.23	3.88	4.60	4.05	3.68	4.45
57	7.89	7.50	8.30	7.72	7.27	8.21	9.79	9.17	10.46	9.50	8.84	10.21	6.33	5.86	6.84	6.13	5.63	6.68
62	11.00	10.35	11.69	10.95	10.25	11.70	13.45	12.65	14.30	13.49	12.64	14.40	8.87	8.13	9.68	8.71	7.95	9.55

References

- [1] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2005.
- [2] S. Ullah and C. F. Finch, "Applications of functional data analysis: a systematic review," *BMC Medical Research Methodology*, vol. 13, article 43, pp. 539–572, 2013.
- [3] R. J. Hyndman and M. S. Ullah, "Robust forecasting of mortality and fertility rates: a functional data approach," *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 4942–4956, 2007.
- [4] S. A. Bashir and J. Estève, "Projecting cancer incidence and mortality using Bayesian age-period-cohort models," *Journal of Epidemiology and Biostatistics*, vol. 6, no. 3, pp. 287–296, 2001.
- [5] E. A. Eisen, J. Bardin, R. Gore, S. R. Woskie, M. F. Hallock, and R. R. Monson, "Exposure-response models based on extended follow-up of a cohort mortality study in the automobile industry," *Scandinavian Journal of Work, Environment and Health*, vol. 27, no. 4, pp. 240–249, 2001.
- [6] L. Hristova, I. Dimova, and M. Ilcheva, "Projected cancer incidence rates in Bulgaria, 1968–2017," *International Journal of Epidemiology*, vol. 26, no. 3, pp. 469–475, 1997.
- [7] E. Negri, C. la Vecchia, F. Levi, A. Randriamiharisoa, A. Decarli, and P. Boyle, "The application of age, period and cohort models to predict Swiss cancer mortality," *Journal of Cancer Research and Clinical Oncology*, vol. 116, no. 2, pp. 207–214, 1990.
- [8] T. Dyba, T. Hakulinen, and L. Päiväranta, "A simple non-linear model in incidence prediction," *Statistics in Medicine*, vol. 16, no. 20, pp. 2297–2309, 1997.
- [9] T. Dyba and T. Hakulinen, "Comparison of different approaches to incidence prediction based on simple interpolation techniques," *Statistics in Medicine*, vol. 19, no. 13, pp. 1741–1752, 2000.
- [10] R. D. Lee and L. Carter, "Modeling and forecasting U. S. mortality," *Journal of the American Statistical Association*, vol. 87, no. 419, pp. 659–671, 1992.
- [11] R. Lee and T. Miller, "Evaluating the performance of Lee-Carter mortality forecasts," pp. 1–17, 2000, <http://www.demog.berkeley.edu>.
- [12] A. E. Renshaw and S. Haberman, "On the forecasting of mortality reduction factors," *Insurance: Mathematics and Economics*, vol. 32, no. 3, pp. 379–401, 2003.
- [13] R. D. Lee, "Modeling and forecasting the time series of US fertility: age distribution, range, and ultimate level," *International Journal of Forecasting*, vol. 9, no. 2, pp. 187–202, 1993.
- [14] S. Tuljapurkar, N. Li, and C. Boe, "A universal pattern of mortality decline in the G7 countries," *Nature*, vol. 405, no. 6788, pp. 789–792, 2000.
- [15] A. Deaton and C. Paxson, "Mortality, income, and income inequality over time in the Britain and the United States," Tech. Rep. 8534, National Bureau of Economic Research, Cambridge, Mass, USA, 2004.
- [16] F. Hollmann, J. Tammany, and J. Kallan, "Methodology and assumptions for the population projections of the United States: 1999 to 2100. Population division, bureau of the census, population division," *International Journal of Forecasting*, vol. 38, 2000.
- [17] B. Erbas, R. J. Hyndman, and D. M. Gertig, "Forecasting age-specific breast cancer mortality using functional data models," *Statistics in Medicine*, vol. 26, no. 2, pp. 458–470, 2007.
- [18] S. N. Wood, "Modelling and smoothing parameter estimation with multiple quadratic penalties," *Journal of the Royal Statistical Society Series B*, vol. 62, no. 2, pp. 413–428, 2000.
- [19] F. Yasmeen, R. J. Hyndman, and B. Erbas, "Forecasting age-related changes in breast cancer mortality among white and black US women: a functional data approach," *Cancer Epidemiology*, vol. 34, no. 5, pp. 542–549, 2010.
- [20] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.
- [21] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK, 1992.
- [22] CBTRUS, *American Cancer Society, Brain Cancer Facts Sheets: Brain and Other Nervous System*. Atlanta, GA, 2010. *Central Brain Tumor Registry of the United States [CBTRUS]: CBTRUS Statistical Report. Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2006*, CBTRUS, 2010.
- [23] SEER, *Surveillance, Epidemiology, and End Results (SEER) Program. SEER Stat Database: Mortality—All COD, Aggregated with State, Total U.S. (1969–2007) <Katrina/Rita Population Adjustment>*, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, June 2010. *Underlying Mortality Data Provided by NCHS*, National Institute of Health, 2010.
- [24] A. J. Coale and E. E. Kisker, "Defects in data on old age mortality in the United States: new procedures for calculating approximately accurate mortality schedules and life tables at the highest ages," *Asian & Pacific Population Forum*, vol. 4, no. 1, pp. 1–31, 1990.
- [25] L. Legler, J. M. Ries, M. Smith et al., "Brain and other central nervous system cancers: recent trends in incidence and mortality," *Journal of National Cancer Institute*, vol. 91, no. 2, pp. 1382–1389, 1999.
- [26] J. P. Neglia, L. L. Robison, M. Stovall et al., "New primary neoplasms of the central nervous system in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study," *Journal of the National Cancer Institute*, vol. 98, no. 21, pp. 1528–1537, 2006.
- [27] S. Horiuchi and J. Wilmoth, "The aging of mortality decline," in *Proceedings of the Annual Meetings of the Population Association of America*, San Francisco, Calif, USA, August 2005.