Coherent Forecasts of US Age-specific Breast Cancer Mortality

Farah Yasmeen
Department of Statistics, University of Karachi, Karachi 75270, Pakistan.

Accepted 6th February, 2015.

Demographers often need to obtain individual forecasts of sub-groups within a population and it is desirable for the disaggregated forecasts to be coherent with the overall forecast. “Coherent” forecasts are non-divergent forecasts of subgroups within a population. In this paper, we intend to obtain coherent forecasts of breast cancer mortality data of black and white women in the United States. This is an application of coherent functional models of Hyndman et al. (2013) on the disaggregation of mortality forecasts by race. On the basis of previous studies, we found that black Americans have higher mortality rates and shorter survival times from breast cancer than white Americans. Here, we assume that the future breast cancer mortality rates of black women will remain higher than those of white women, for all age groups. We first describe the concept of coherence in the context of cause-specific mortality and discuss some problems with using independent functional time series models for the two races. Then, we apply the coherent functional model to the breast cancer mortality data. An empirical comparison of the independent and coherent models based on the breast cancer mortality data has been made. The purpose here is to see the performance of coherent forecasting models in the presence of disparity among the mortality rates of these two groups of American women, namely whites and blacks and to find whether the coherency is achieved in breast cancer mortality forecasting by this application.

Keywords: Forecasting, functional data analysis, coherent forecasts, breast cancer mortality, racial and ethnic disparities, trends.

1. INTRODUCTION

In the United States, the higher mortality rates and lower incidence rates in black women than white women are largely unexplained (Chelbowski et al. 2005). In the United States, many studies have shown that the trends in breast cancer mortality rates for white women are not the same as those for black women (see Ghafoor et al. 2003, Tarone et al. 1997, Lacey & LA 2002).

Yasmeen et al. (2010) have shown that there are major differences in the mortality rates of the two races by age. They applied the functional time series models (Hyndman & Ullah 2007) to breast cancer mortality rates among white and black women in the United States. In this paper, we intend to obtain coherent forecasts of breast cancer mortality data for black and white women in the United States.

This is an application of coherent functional models of Hyndman et al. (2013) on the disaggregation of mortality forecasts by race. The purpose here is to see the performance of the coherent forecasting model in presence of disparity
among the mortality rates of these two groups of American women, namely whites and blacks.

This paper is organized as follows. In section 2, we describe the concept of coherence and section 3 applies the coherent functional model to the breast cancer mortality data. Section 4 is an empirical comparison of the independent and coherent models based on the breast cancer mortality data. In section 5, we discuss the results obtained from the application of the new methods, and finally, some conclusions are drawn in section 6.

Data

We will consider the annual age-specific breast cancer mortality rate data obtained from the Surveillance, Epidemiology and End Results (SEER) program in the United States (see www.seer.cancer.gov). We will use the mortality data for the women of the two races (whites and blacks) from 1969 through 2010 in eight 5-year age groups: 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79 and 80–84.

2. COHERENT FORECASTS FOR THE TWO RACES

Demographers often need to obtain individual forecasts of subgroups within a population and it is desirable for the disaggregated forecasts to be coherent with the overall forecast. "Coherent" forecasts are non-divergent forecasts of subgroups within a population. Here, we can define the concept of coherence as "the mortality rates of one race will remain higher than the rates of another race, across all age groups". Several recent publications have shown an overall decline in breast cancer mortality in the 1990s (for example, see Jemal et al. 2002, 2004), resulting in a disparity between the rates of whites and blacks. A better definition of coherence can be defined in the following way:

If \( ft(x) \) is a function in one group and \( gt(x) \) is a function for the same time in the other group, then we require

\[
\lim_{t \to \infty} \sup_{x} |ft(x) - gt(x)| < \infty
\]

One way to achieve that is for \( dt(x) = ft(x) - gt(x) \) to be a stationary functional process. This is much more precise in what is meant by 'not diverge in the long term'. The definition given above does not cover all types of non-divergence.

Figure 1 shows twenty-year forecasts of breast cancer mortality rates for white and black women using independent functional time series models. It is evident that the use of independent models provides forecasts that are not coherent according to either of the assumptions given above. Similarly, the independent forecasts show that the difference between the mortality rates of the two races is increasing. Within the same population in the US, the mortality rates are decreasing for whites, but for blacks, there is an indication of stability or very little improvement over the next twenty years, which is also against our concept of coherence.

3. APPLICATION OF THE COHERENT FUNCTIONAL MODEL

Hyndman et al. (2013) proposed the coherent functional model that provides forecasts of the mortality rates of two or more groups within a given population by simultaneously modeling the geometric mean of the age-specific mortality rates and the ratio of the group-specific rates to the geometric mean of the mortality rates of the population as a whole.

3.1. Coherent Functional Model for White and Black Women

Let \( mt,W(x) \) denote the breast cancer mortality rate for whites of age \( x \) in year \( t \), \( t = 1, \ldots, n \). We model the log mortality rate, \( yt,W(x) = \log[mt,W(x)] \), and assume that there is an underlying smooth function \( ft,W(x) \) that we are observing with error. Thus,

\[
yt,W(x) = \log[ft,W(x)] + \epsilon_t,W,i,
\]

where \( xi \) is the centre of age group \( i (i = 1, \ldots, p) \), \( \epsilon_t,W,i \) is an independent and identically distributed standard normal random variable and \( \epsilon_t,W,i \) allows the amount of noise to vary with age \( x \). Analogous notation can be used for blacks.

Define the square roots of the products and ratios of the smoothed rates for each race group:

\[
pt(x) = ft,W(x)/ft,B(x) \quad \text{and} \quad rt(x) = ft,W(x)/ft,B(x).
\]

3.2. Results

In applying the coherent functional model to the breast cancer mortality data of white and black women, the first step is to obtain the product component \( pt(x) \) and ratio component \( rt(x) \) using equation (2). Figures 2 and 3 depict these components of the smoothed log mortality rates during the years 1969–2010. The product component represents the overall effect of the whole population (irrespective of groupings), and it shows that overall, the mortality rates for women of both races increased with age. Also, during the study period, the rates of women below 55 years of age decreased, whereas for women over 55 years in both races, they first increased and then decreased in recent years. For women older than 75 years, the mortality rate increased continually, and the decline in rates appeared in the most recent years.

The ratio component shows the square root of the ratio of the death rates of whites to the death rates of blacks, as defined in equation (2), as displayed in Figure 3. This figure suggests that the white to black mortality ratio was more than 1.0 in early years, especially for women over 55 years of age. However, it decreased continually since that time, and in most recent years, the ratio has consistently been less than 1 for all age groups. This suggests that white women have lower mortality rates than black women for all ages during the study period.

The next step is to apply the basic functional time series model (Hyndman & Ullah 2007) to the product and ratio components. Figures 4 and 5 show the basis functions and the respective coefficients obtained by applying the functional principal components decomposition to the product and ratio curves, along with twenty-year forecasts of the time series coefficients. These plots represent the various sources of variation in the two components. From Figure 4, it is clear that the mean function of the product model is an increasing function of age. The first basis function of the product component shows that the breast cancer mortality rates for the two groups of women (whites and blacks) are decreasing for younger women.

This decline started in 1990. Similarly, the second basis function shows that the rates for older women (ages 70–80 years) were initially increasing (since 1969) and then declining in 1990. However, this decline was more rapid in younger women than in older women. We use ARIMA models to forecast the product component coefficient, using automatic
**Figure 1**: Twenty-year breast cancer mortality forecasts (2011–2030) for white and black women using independent functional time series models. The curves are ordered in time using rainbow colors. The earliest year (from 2011) are shown in red, with the latest years (from 2030) shown in purple.

**Figure 2**: Log of the product component of the coherent functional model for the breast cancer mortality data of white and black women (1969–2010). The curves are ordered in time using rainbow colors, with the later years in violet.

**Figure 3**: White to black breast cancer mortality ratio (1969–2010). The curves are ordered in time using rainbow colors, with the later years in violet.
Figure 4: The first three basis functions and the corresponding coefficients of the product model for the breast cancer mortality data, together with twenty-year forecasts from ARIMA models.

Figure 5: The first three basis functions and corresponding coefficients of the ratio model for the breast cancer mortality data, together with twenty-year forecasts obtained from stationary ARFIMA \((p,d,q)\) \((-0.5 < d < 0.5)\) models.
Figure 6: 80% prediction intervals of one-year forecasts (2011) of US Breast Cancer Mortality using Coherent Functional models. Red color is used for whites and blue for blacks.

Figure 7: 80% prediction intervals of 20-year forecasts (2030) of US Breast Cancer Mortality using Coherent Functional models. Red color is used for whites and blue for blacks.

Figure 8: Product-ratio forecasts (2011-2030) of US Black and White Breast Cancer Mortality using Coherent Functional models.
If we compare the figures 4 and 5, it can be observed that the coefficients of product models are non-stationary with slightly wide prediction intervals. Whereas the coefficients of ratio models are stationary with narrow prediction intervals. Hence, the stationarity condition makes the forecasts of the two groups to be non-divergent.

Figures 6 and 7 show twenty-year forecasts of age-specific breast cancer mortality for white and black women separately. These graphs show a decline in breast cancer mortality rates for both black and white women over the next twenty years. This decline is steeper among older white women (60 years and above) and slower among younger white women (under 55 years of age). For black women, future mortality rates are expected to decline substantially for women aged between 60 and 75 years, and relatively slowly for women of other ages.

Table 1 provide 20-year forecasts (2011–2030) of breast cancer mortality rates for the two races by age group. These forecasts, together with 80% prediction intervals of mortality rates for selected ages in 2011 and 2030, are shown in this

Table 1: Forecasts and 80% prediction intervals (PI) of mortality rates (per 100,000 women) obtained from the coherent functional model for selected ages in 2011 and 2030

<table>
<thead>
<tr>
<th>Age</th>
<th>White women</th>
<th>Black women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast 2011</td>
<td>Forecast 2030</td>
</tr>
<tr>
<td>Mean</td>
<td>80% PI</td>
<td>Mean</td>
</tr>
<tr>
<td>65</td>
<td>31.90</td>
<td>[29.48, 34.51]</td>
</tr>
<tr>
<td>70</td>
<td>38.43</td>
<td>[35.37, 41.74]</td>
</tr>
<tr>
<td>75</td>
<td>49.61</td>
<td>[45.35, 54.26]</td>
</tr>
<tr>
<td>80</td>
<td>78.86</td>
<td>[72.11, 86.24]</td>
</tr>
</tbody>
</table>

forecasting algorithm of Hyndman & Khandakar (2008) to select the best possible ARIMA model.

The mean function in Figure 5 shows that on average, the white to black mortality ratio increases with age, suggesting that mortality rates for older women for both races are similar and any differences are predominantly at younger ages. The first basis function of the ratio component shows that most of the variation in the white to black mortality ratio is in the age group 50-60, with corresponding rates decreasing from 1969 and stabilizing from 1995.

For forecasting the ratio coefficients, we use stationary ARFIMA (p, d, q) (−0.5 < d < 0.5) models and obtain forecasts using the algorithm developed in Hyndman et al. (2013). The twenty-year forecasts of the first ratio coefficient show that the white to black mortality ratio may remain constant in the future. This implies that the forecasts for the entire mortality curves for white and black women will not diverge in the long run. The other basis functions and coefficients have little effect on the ratio forecast, as the forecast values of the coefficients are nearly zero.

Figure 9: MSFE for whites and blacks using independent and coherent models. Solid lines are used for independent and dotted lines for coherent model.
It has been found that coherent models have about the same pattern for age-specific forecasts as the results obtained in Table 1 of Yasmeen et al. (2010) using independent models. However, the 80% prediction intervals from the new approach are much tighter.

4. EMPIRICAL COMPARISON OF INDEPENDENT AND COHERENT MODELS

In this section, we will compare the results obtained from the application of independent and coherent functional models to the breast cancer mortality data of US women.

Figure 8 shows comparisons between the independent functional time series model and coherent model for the two races. These figures suggest that the new method performed better than the existing method for white women at all forecast horizons. The coherent model performs excellently in reducing the MISFE, especially for the longer forecast horizons. However, for black women, the new methods provide approximate levels of equal precision, and their performances are relatively lower than that of the independent model at all forecast horizons. One possible reason for this is that the first basis function of the product component of the coherent model (which is more responsible for the overall variation in the mortality rates of all women) is more similar to the first basis function of the independent functional time series model of whites, than to that for blacks.

5. DISCUSSION

Here, we applied newly developed functional time series models to the age-specific black and white breast cancer mortality data in the United States. These new models include coherent functional models based on the products and ratios of the mortality rates. We have shown that, on average, black women have had higher breast cancer mortality rates than their white counterparts in all age-groups since 1990, when the rates started to decline possibly due to medical interventions and early detection activities becoming more prevalent among white populations. Our forecasts of future mortality rates show that these differences are expected to be maintained over the next twenty years.

Our analysis suggests that the first common basis function, which is the major source of variation and most responsible for the changes, corresponds to younger age groups for women of both races. All breast cancer mortality rates are expected to decline in future. However, the rate of decline in whites is expected to be greater for women aged between 60-75 years, and relatively small in other age groups. For black women, the plot of the forecasts shows that the rates will decline very rapidly for women aged 60-70 years and relatively slowly for other ages. However, black women are expected to have higher mortality rates than whites over the next twenty years for all age-groups.

The results obtained here are consistent with the findings of Jemal et al. (2007) and Dawood et al. (2008) and the results obtained in Yasmeen et al. (2010), that the overall breast cancer mortality rates are decreasing. This decline can be attributed to the increased participation of both groups of women in screening programs. Ansell et al. (2009) reported a study attempting to reduce the disparity between black and white mortality rates in Chicago. They observed that the black and white cancer mortality rates were similar in 1980. By the late 1990's, a substantial disparity was present, and by 2005, the black breast cancer mortality rate (age-adjusted) was about 116% higher than the white rate. The authors found significant barriers to access to high quality mammography and treatments that could be contributing to the differences between the mortality rates of women of the two races.

In a population-based analysis attempting to determine the underlying causes of the black-white mortality disparity, Menashe et al. (2009) found that, black to white mortality rate ratio was greater than one between 1990 and 2004, and it had been widening ever since. The age-standardized breast cancer mortality rates fell from 30 to 22 per 100,000 persons in whites and from 36 to 29 per 100,00 in blacks. Baquet et al. (2008) performed an analysis on three age-groups: < 40, 40-49, and ≥ 50, on black/white mortality data for the period 1995-2004. They found that, black women were significantly more likely to be diagnosed with regional or distant disease than whites, and have a lower five-year survival rate. Relative to white women, black women often appear with advanced-stage breast cancer, which may explain the higher observed mortality among blacks (see Deshpande et al. 2009).

Our analysis has also shown that the major differences between the mortality rates of the two races, white and blacks, are in the age group 50-70 years, which might be the target age group for screening programmes in the United States. The findings of our research are consistent with the results obtained by other studies such as those of Carney et al. (2003) and Norman et al. (2007), who suggested that women aged 40-64 years can benefit from mammography in reducing the risk of death from breast cancer. Mammography screening is more effective in older women (aged 50-64 years) than in younger women (aged 40-49). The screening efficacy is usually greater in postmenopausal women, as the breast density affects mammographic sensitivity (Carney et al. 2003) and the density decreases after menopause (Stone et al. 2003).

However, some of the results obtained are slightly different. The study by Tyczynski et al. (2006) reported that black women are more likely to be without health insurance than whites. Among older women with Medicaid, more blacks do not have supplemental health insurance. They also found that these women tend to report difficulties in gaining access to health care than white women, and postmenopausal black American women do not benefit from the overall decline in the breast cancer mortality rates. They analyzed the data for the state of Ohio and used five 10-years age groups: 30-39, 40-49, 50-59, 60-75, and over 75 years. In contrast, we performed an analysis for all age groups of white and black women and found that future mortality rates are declining for both races, with the mortality rates being higher for blacks than for whites in all age groups. We also performed an age-specific analysis of the breast cancer mortality among black and white women and found that there are differences in the age group 50-69 years, and these differences are expected to continue into the future.

Marbella & Layde (2001) calculated the ratios of black to white breast cancer mortality rates for women from 1979 to 1996 in five-year age groups: 25 – 29, 30 – 34, . . . , 85+. They found that during the study period, black women 45 years and younger had consistently higher breast cancer mortality rates than white women of the same age. From 1993 to 1996, the rates for black women aged 40–50 years, were 1.5% higher than those for whites. The authors found
that between 1989 and 1992 and between 1993 and 1996, white women of all ages showed consistently decreasing patterns, whereas the rates for blacks had no consistent patterns.

In contrast, we plotted the ratio of white to black mortality rates (see Figure 6). The mean function in Figure 8 shows that on average, this ratio increases with age, and hence, the differences in the rates of two races are mostly at young ages. The first basis function of the ratio component shows that most of the variation in the white to black mortality ratio is in the age group 50-60 years, where the corresponding rates decreased from 1969, then started to stabilize from 1995. We assumed that this ratio will remain constant (stationary) over the next twenty years, so that the resulting mortality forecasts for the two races will not diverge.

6. CONCLUSION

In this paper, we applied the newly developed functional time series models of Hyndman et al. (2013) to the age-specific black and white breast cancer mortality data of the United States. These models include coherent functional models based on the products and ratios of the mortality rates. From the application in Yasmeen et al. (2010), it was found that the use of independent functional time series models on the two races of US women does not guarantee that the resulting mortality rate forecasts will be non-divergent. We found that the newly developed methods provide breast cancer mortality forecasts that are both coherent with the historical trends in the two races and more realistic.

On the basis of the mean integrated square forecast error (MISE), it is found that for the two groups (combined), the forecasts from the new methods are better than the forecasts obtained from the independent models. However, the coherent functional model has a tremendous performance in forecasting breast mortality rates for whites, whereas the independent functional time series model is appeared to be better for blacks. From Figure 9, one can say that the overall gain in precision for the two groups (combined) is higher using the coherent models, as compared to independent model.

Breast cancer mortality forecasts with 80% prediction intervals are also obtained for both races of US women. These forecasts show that the breast cancer mortality rates are expected to decline in the future, with the mortality rates of blacks expected to remain higher than those of whites for all age groups. Our analysis suggests that the disparity between the mortality rates of the two races in the two races is expected to continue over the next twenty years, which will have implications for breast screening management programs.

REFERENCES


