

Statistical Data

Using Functional Data Analysis Models to Estimate Future Time Trends in Age-Specific Breast Cancer Mortality for the United States and England–Wales

Bircan Erbas¹, Muhammed Akram², Dorota M Gertig³, Dallas English^{4,5}, John L. Hopper⁵, Anne M Kavanagh⁶, and Rob Hyndman²

¹School of Public Health, La Trobe University, Bundoora, 3086 Australia

²Business & Economic Forecasting Unit, Monash University, Clayton, 3800, Australia

³Victoria Cytology Service Inc, Carlton, 3053 Australia

⁴Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton 3053 Australia

⁵Centre for MEGA Epidemiology, The University of Melbourne, Parkville 3053 Australia

⁶Key Centre for Women's Health in Society, School of Population Health, The University of Melbourne, Parkville, 3053 Australia

Received May 12, 2009; accepted September 3, 2009; released online February 6, 2010

ABSTRACT

Background: Mortality/incidence predictions are used for allocating public health resources and should accurately reflect age-related changes through time. We present a new forecasting model for estimating future trends in age-related breast cancer mortality for the United States and England–Wales.

Methods: We used functional data analysis techniques both to model breast cancer mortality-age relationships in the United States from 1950 through 2001 and England–Wales from 1950 through 2003 and to estimate 20-year predictions using a new forecasting method.

Results: In the United States, trends for women aged 45 to 54 years have continued to decline since 1980. In contrast, trends in women aged 60 to 84 years increased in the 1980s and declined in the 1990s. For England–Wales, trends for women aged 45 to 74 years slightly increased before 1980, but declined thereafter. The greatest age-related changes for both regions were during the 1990s. For both the United States and England–Wales, trends are expected to decline and then stabilize, with the greatest decline in women aged 60 to 70 years. Forecasts suggest relatively stable trends for women older than 75 years.

Conclusions: Prediction of age-related changes in mortality/incidence can be used for planning and targeting programs for specific age groups. Currently, these models are being extended to incorporate other variables that may influence age-related changes in mortality/incidence trends. In their current form, these models will be most useful for modeling and projecting future trends of diseases for which there has been very little advancement in treatment and minimal cohort effects (eg. lethal cancers).

Key words: breast cancer; forecasting; functional-data-analysis models; mortality trends

INTRODUCTION

Breast cancer is the most commonly diagnosed cancer in women worldwide.¹ In the United States and United Kingdom, breast cancer is the second highest cause of cancer death in women.^{2,3} Therefore, accurate projections of mortality/incidence from breast cancer are important for planning future public health policy and resource allocation. In particular, accurate age-specific projections of mortality from breast cancer are essential for assessing cancer control programs such as mammographic screening. Screening, combined with improvements in treatment options,

influences the mortality and incidence patterns for women of different ages.^{4,5} Because trends in breast cancer mortality and incidence vary substantially with age, it is important that predictions accurately consider and reflect these variations.

There are noticeable differences in breast cancer mortality patterns over time by age across different countries. In the United Kingdom, breast cancer mortality has decreased substantially for women between the ages of 55 and 69 years, as compared with those aged 50 to 54 years.⁶ Conversely, in the United States, breast cancer mortality for white women has declined more rapidly for those younger than 50 years than for those who are older.⁵ A decrease in

Address for correspondence: Dr. Bircan Erbas, La Trobe University, Bundoora, Melbourne 3086, Australia (e-mail: bircan.eras@gmail.com).

Copyright © 2010 by the Japan Epidemiological Association

mortality for women aged 30 to 49 years has been observed in a number of European countries that lack an organized nationwide screening program.⁴ Mortality among women older than 65 years has continued to increase in many countries, regardless of screening or advances in treatment.^{7,8}

Differences in age-related mortality trends between countries may reflect differences in mammographic screening policies and age-related differences in the uptake of treatment options (such as tamoxifen). To take one example, there is triennial screening of 50- to 70-year-olds in the United Kingdom,⁹ while the United States has no organized screening program and requires referral by a medical practitioner.¹⁰ Other factors that may contribute to these differences are hormone replacement therapy use and oral contraceptive use. When modeling incidence/mortality trends, these factors are known as both period and cohort effects. Given these apparent age-related differences, it may be misleading to estimate future breast cancer mortality patterns without accounting for age-related time trends in mortality.

In this study, we used a recently developed forecasting method¹¹ to (1) compare the time trends of age-specific breast cancer mortality for the United States and England–Wales and (2) predict future rates of age-specific breast cancer mortality for the United States and England–Wales. The forecasting method we use predicts the entire age-mortality relationship through time, and does not simply rely on the most recent data. The good forecasting performance of the present models in other contexts¹² suggests that the models are likely to increase the accuracy of predictions in the present context as well. To the best of our knowledge, this is the first study to incorporate the effect of age-specific trends over time on breast cancer mortality when estimating future trends of age-specific breast cancer mortality in England–Wales and the United States.

METHODS

Data

Annual age-specific breast cancer mortality data for England and Wales from 1950 through 2003 were obtained from the Office of the National Statistics.³ US mortality data from 1950 through 2001 were extracted from the World Health Organization mortality database.¹³ Data for breast cancer are designated by ICD-6 and -7 code 170, ICD-8 and -9 code 174, and ICD-10 code C50. The data are available in 5-year age groups: 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, and 80–84 years.

Statistical analysis

Breast cancer mortality was observed annually as a function of age, defined as the midpoint of the age groups. For each year, we plotted the age and mortality associations, which we refer to as mortality-age curves. We then took the log of the mortality rate (our outcome) at each midpoint of age for each year and used functional data analysis (FDA) techniques¹⁴ to

model these annual mortality-age curves collectively as a functional time series. In these models we first assumed an underlying smooth function that we are observing with error (see the Appendix for details). We then used nonparametric regression techniques to estimate the smooth curves.^{15,16}

Next, we took these smooth curves as our functional observations and fit functional data analysis models. We followed the estimation procedure of Hyndman and Ullah¹¹ and applied functional principal components decomposition¹⁷ to the smooth curves, because this approach produces a small number of basis functions, enables informative interpretations, and yields coefficients that are uncorrelated with each other.

To predict future mortality, we forecast each coefficient in the models using a univariate time series model. We multiplied these forecasts with the basis functions in the FDA models, resulting in forecasts of mortality age curves through time. We then used exponential smoothing state space models to compute the forecasts¹⁸ and construct prediction intervals around our predictions.¹⁹ We used the Mean Integrated Squared Forecasting Error (see Appendix) to evaluate the accuracy of the estimated predictions of future mortalities. For the US and England–Wales mortality data we estimated 20-year predictions using exponential smoothing state space models with damping.¹⁸ All statistical analyses were performed using R version 2.3.1.

RESULTS

Figure 1 displays the observed breast cancer mortality time trends by age for the United States (per 100 000 women) from 1959 through 2001 (left) and for England–Wales (per 100 000 women) from 1950 through 2003 (right). For the United States, mortality trends for middle-aged women (45–54 years) have continued to decline since 1980. Mortality trends for women between 60 and 84 years of age increased in the 1980s and subsequently declined in the 1990s. The observed pattern of mortality is similar for England–Wales. Mortality trends for women aged 45 to 74 years slightly increased between 1950 and 1980, but declined thereafter. Mortality trends for women aged 75 to 79 years and 80 to 84 years fluctuated during the study period. However, as was the case for the other age groups, there was an overall decline in mortality during the 1990s.

Under our model, an adequate fit (as determined by the MISE) was a functional regression model with 2 basis functions for the US data; for England–Wales, it was a functional regression model with 3 basis functions. The first basis function accounts for 68.4% of variation around the mean log mortality curve for the United States and 71.5% of variation around the mean log mortality curve for England–Wales. For England–Wales, β_1 showed an increase in mortality for all age groups between 1950 to 1980, followed by a rapid decline until 2000. Similar interpretations can be made for the United States.

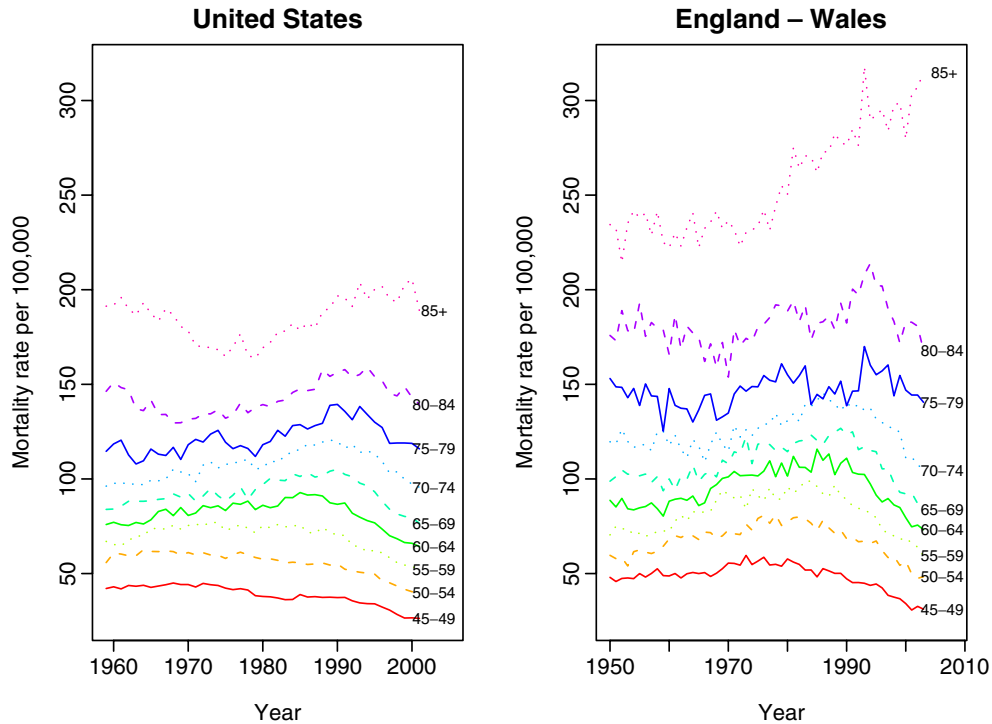


Figure 1. Observed breast cancer mortality trends by age group for the United States (left) and England–Wales (right).

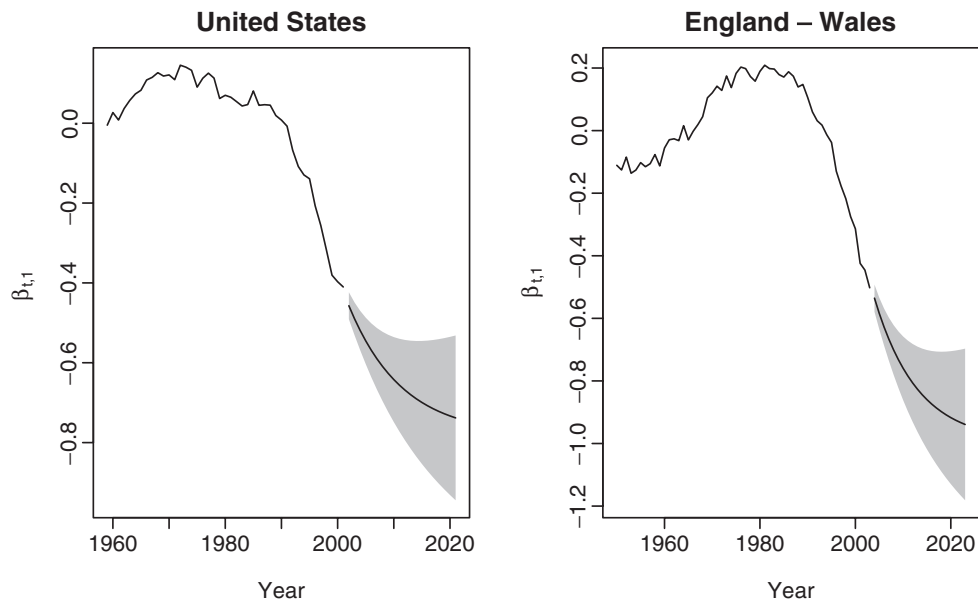


Figure 2. Twenty-year mortality predictions for the United States (left) and England–Wales (right) using a damped trend exponential smoothing model. The y-axis represents the estimated coefficient of the first basis function. The shaded region gives the 80% prediction interval.

Twenty-year projections of the first basis function, which controls the overall change in trend of breast cancer mortality, are shown in Figure 2, along with 80% prediction intervals, for (1) the United States from 2002 through 2021 and (2) England–Wales from 2004 through 2023. The y-axis represents the coefficients associated with the first basis function, ie, the overall change in trend of breast cancer

mortality. Predictions from the models suggest that overall crude mortality rates for both countries will decline more slowly than during the 1990s. These predictions assume no changes or advances in treatment.

Figures 3a and 3b display 20-year predictions of age-specific breast cancer mortality for both regions. Mortality trends are expected to decline for all women, with the greatest

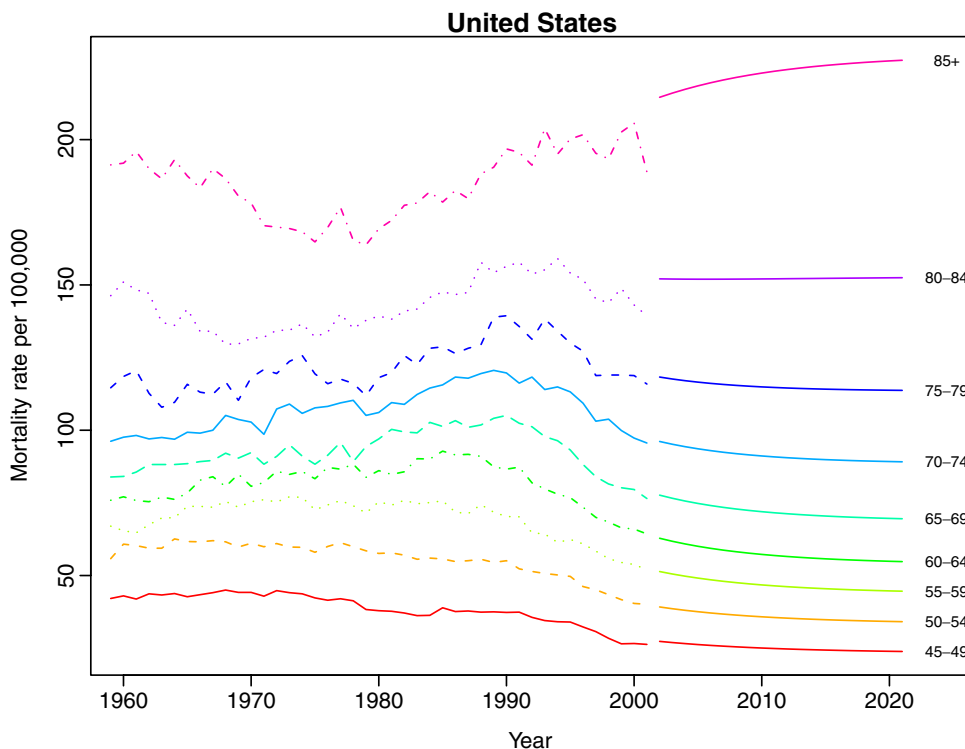


Figure 3a. Estimated 20-year predictions of age-specific breast cancer mortality for the United States.

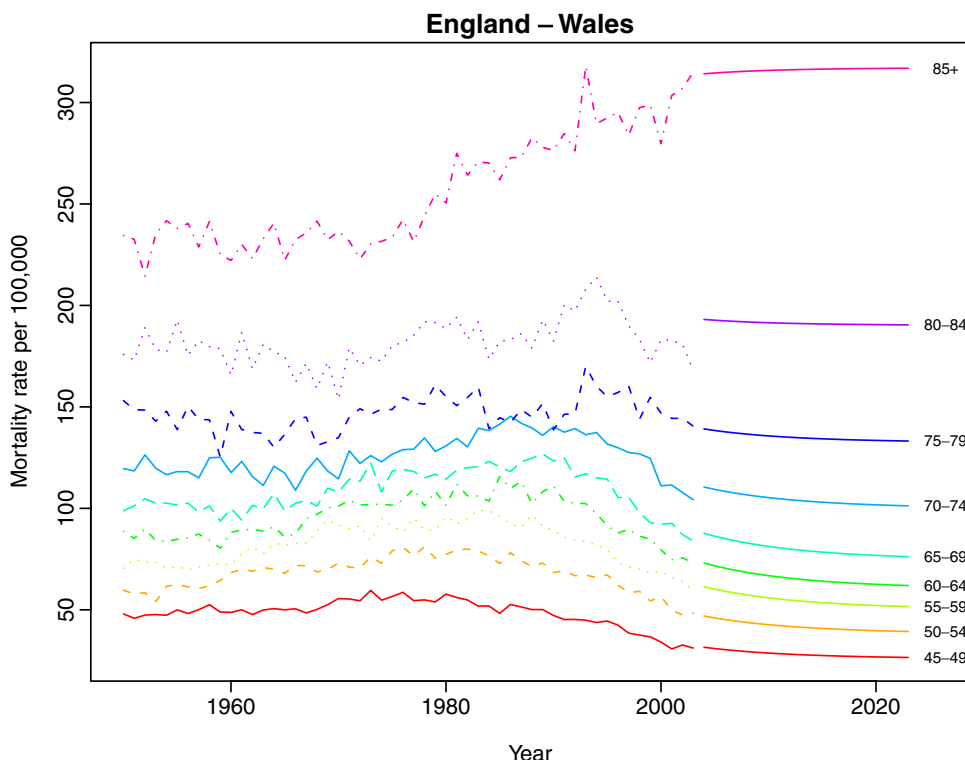


Figure 3b. Estimated 20-year predictions of age-specific breast cancer mortality for England–Wales.

decline among women aged 60 to 70 years, whereas estimated predictions suggest relatively stable trends for women older than 75 years.

To evaluate the accuracy of the predictions, we estimated 1-, 10-, and 20-year age-specific mortality predictions with

80% prediction intervals for both regions (Figure 4). The estimated predictions had very narrow prediction bands. For example, 20-year predictions for 60-year-old women in England–Wales had an 80% error margin of less than 10 deaths per 100 000 women.

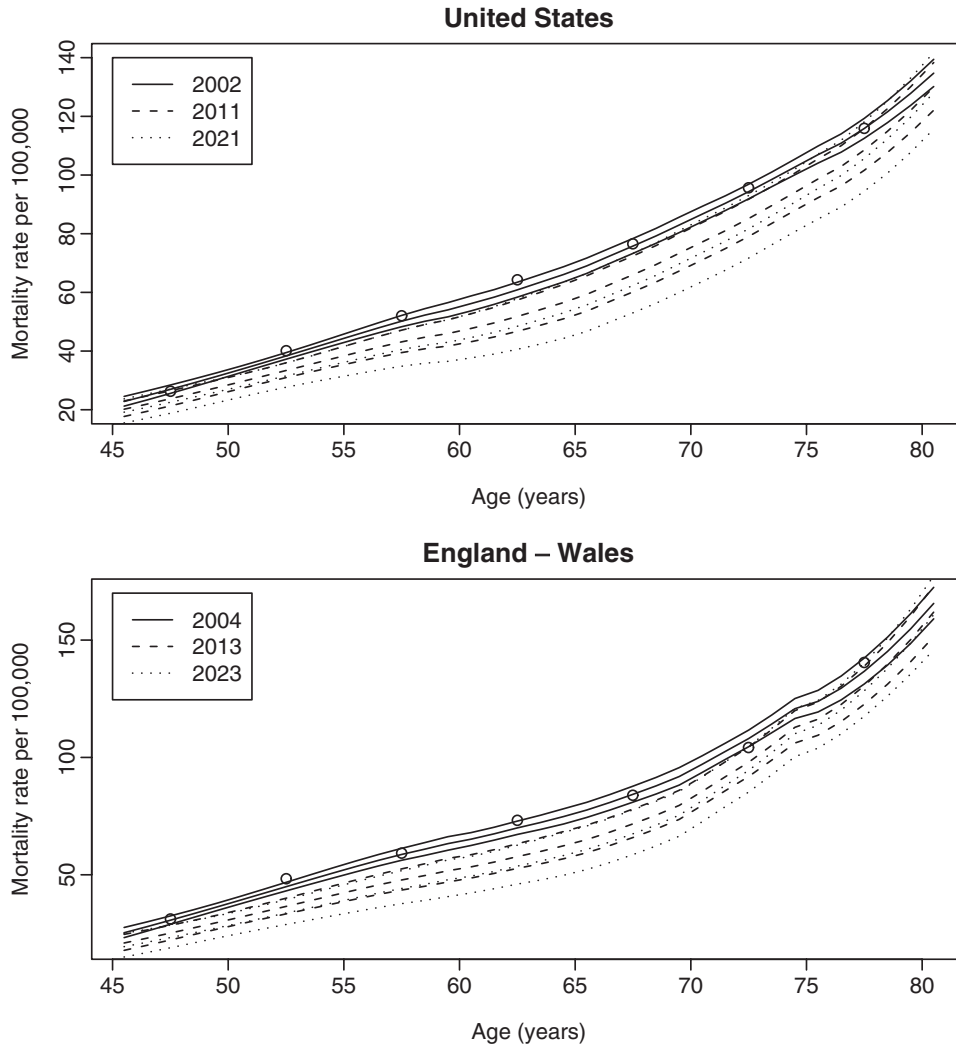


Figure 4. Estimated 1-, 10-, and 20-year age-specific predictions for the United States (top) and England–Wales (bottom). The circles represent actual values.

We examined the residual of the functional fits using image plots (data not shown). These images showed no evidence of lack of fit and suggest that the few remaining birth cohort trends are of no concern.

DISCUSSION

Using an innovative forecasting method, our 20-year projections suggest a continuing decline in breast cancer mortality for both England–Wales and the United States. There has been considerable debate regarding the primary factors responsible for the decline in mortality observed in England–Wales, and to a lesser extent in the United States.^{20,21} In England–Wales, the decline in mortality since the early 1990s has often been attributed to better treatment practices and the widespread use of tamoxifen.²⁰ At the end of the 1990s and more recently, screening is also believed to have played a role, but there are doubts as to whether it had a major impact on mortality in the early 1990s, when the sharpest decline occurred.^{22,23} Since 1988, the United Kingdom has

had an organized mammographic screening program in place for women aged 50 to 70 years,²⁴ whereas in the United States screening is more ad-hoc and is recommended for women over 40 years of age.²¹ It is therefore possible that screening contributed to the greater decline in women older than 75 years in the United Kingdom, but not in the United States. Other factors, such as the rapid increase in hormone therapy use in the 1990s and the subsequent decline in its use after the release of the Women’s Health Initiative trial findings,²⁵ may have contributed to the continuing decline in breast cancer mortality for England–Wales. The reduction observed in the United States has been attributed to early detection through screening, combined with the use of adjuvant chemotherapy and tamoxifen for patients with all stages of breast cancer.²⁶

This study has a number of important strengths. First, we have presented an alternative modeling approach to the classic age-period-cohort (APC) models. Here, mortality rates are regarded as smooth functions of age, and the shape of the mortality-age relationship is allowed to change over time. Our modeling and forecasting approach is appealing because of the

few assumptions required and the visual character of the projections. Second, unlike most other studies, we estimated prediction intervals for future mortality rates. Prediction intervals are necessary to accommodate uncertainty in the mathematical structure of the model and variation in the future mortality rate.²⁷ The prediction intervals from the functional forecasting models are narrow for both the overall and age-specific breast cancer mortality rates in England–Wales and the United States, suggesting that the models have captured the stochastic and dynamic properties of the data. As expected, prediction intervals widened as the forecast horizon increased.

A number of limitations should be considered when interpreting the long-term mortality trends reported here. First, changes in coding practices, the accuracy of death reporting throughout the period of data collection, revisions of ICD codes, and the combining of subsites into one major site may affect our interpretation of cancer trends²⁸; however, for breast cancer, the general consensus is that the consistency of the data has been reasonable.²⁹

Second, birth cohort effects due to changes in the underlying risk factors were not included in the models and these may alter the shape of the age-mortality distribution.²⁷ Birth cohort trends for US women born from 1880 through 1915 would have affected the estimates and resulted in higher-than-expected projections for older women. At present, the decreasing birth cohort trends for baby boomers have not had their full impact on the decline in future mortality rates in the United States. In a study of the mortality benefits of screening in England–Wales,³⁰ birth cohort effects were similar for all ages before the introduction of screening. The smoothing process used in our modeling may reduce the variation attributable to birth cohort effects. Furthermore, any remaining birth cohort effects will be saturated in patterns of variation over time. Nevertheless, birth cohort trends are an important aspect of modeling mortality/incidence trends and we are currently developing models which incorporate cohort effects and other variables that may influence age-related changes in mortality (or incidence).

In summary, we presented a new modeling and forecasting technique to model and estimate future trends in breast cancer mortality. At present, much of the modeling and prediction of mortality trends uses APC models. Here, we presented an alternative framework, which can predict entire age-mortality curves for each period, to estimate predictions, thereby enhancing the accuracy of the predictions. In their current form, these models will be most useful for modeling and projecting the future trends of diseases for which there has been very little advancement in treatment and minimal birth cohort effects, such as cancers of the pancreas and brain.

APPENDIX

We used functional data analysis techniques¹⁴ to model annual mortality-age curves collectively as a functional time series.

Specifically, let $y_t(x)$ denote the log mortality at age x for year t . We assume that there is an underlying smooth function $f_t(x)$ that we are observing with error. Thus, we observe the functional time series $\{x_i, y_t(x_i)\}$, $t = 1, \dots, n$, $i = 1, \dots, p$, where

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad (1)$$

where $\{\varepsilon_{t,i}\}$ are independent and identically distributed random variables with zero mean and unit variance, and $\sigma_t(x_i)$ describes how the amount of error varies with x .

The smoothed curves $\{f_t(x)\}$ are estimated using nonparametric regression techniques, such as penalized regression splines¹⁵ and Loess curves.¹⁶ The n smoothed curves are our functional observations, $\{f_t(x)\}$, where $x_1 < x < x_p$ and $t = 1, \dots, n$. Erbas, Hyndman, and Gertig¹⁹ and Hyndman and Ullah¹¹ proposed the following model for the smoothed curves:

$$f_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x), \quad (2)$$

where $\mu(x)$ is a measure of the mean log mortality across years, $\{\phi_k(x)\}$ is a set of orthonormal basis functions, and each $\beta_{t,k}$ is a univariate time series. We follow their estimation procedure by computing $\{\phi_k(x)\}$ using functional principal components decomposition¹⁷ applied to the smoothed curves $\{f_t(x)\}$, as this approach produces a small number of basis functions, enables informative interpretations, and yields coefficients that are uncorrelated with each other. For a given value of K , we choose a model with basis functions $\{\phi_k(x)\}$ that minimize the mean integrated squared error (MISE): $= \frac{1}{n} \sum_{t=1}^n \int e_t^2(x) dx$

To make predictions of future mortality rates, we forecast each coefficient $\{\beta_{t,k}\}$ using a univariate time series model. We multiply these projections by the basis functions, resulting in projections of mortality curves $f_{n+h}(x)$, $h = 1, 2, \dots$. We then use exponential smoothing state space models to compute the forecast¹⁸ and construct prediction intervals around our predictions.¹⁹ The Mean Integrated Squared Forecasting Error: $(\text{MISFE}(h)) = \frac{1}{n-m+1} \sum_{t=m}^n \int [y_{t+h}(x) - \hat{f}_{t,h}(x)]^2 dx$ is used to evaluate the accuracy of the estimated predictions of future mortalities.

REFERENCES

1. Parkin DM, Pisani P, Ferlay J. Global Cancer statistics. *CA Cancer J Clin.* 1999;49:33–64, 1.
2. Breast Cancer Facts and Figures. American Cancer Society, Inc., 250 Williams St., Atlanta, GA 30303-1002. (404) 320-3333. Available from <http://www.cancer.org>.
3. Cancer stats. Cancer Research UK is a registered charity No. 1089464. Registered as a company limited by guarantee in England & Wales No. 4325234. Registered address 61 Lincoln's Inn Fields, London, WC2A 3PX. Available from: <http://info.cancerresearchuk.org/cancerstats/types/breast/>.
4. Tyczynski JE, Plesko I, Aareleid T, Primic-Zakelj M, Dalmas M,

- Kurtinaitis J, et al. Breast Cancer mortality patterns and time trends in 10 new EU member states: Mortality declining in young women, but still increasing in the elderly. *Int J Cancer*. 2004;112:1056–64.
5. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, et al. Trends in breast cancer by race and ethnicity: update 2006. *CA Cancer J Clin*. 2006;56:168–83.
 6. Jatoi I, Miller AB. Why is breast-cancer mortality declining? *Lancet Oncol*. 2003;4:251–4 Review.
 7. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. *J Natl Cancer Inst Monogr*. 2006;36:19–25.
 8. Clèries R, Ribes J, Esteban L, Martínez JM, Borràs JM. Time trends of breast cancer mortality in Spain during the period 1977–2001 and Bayesian approach for projections during 2002–2016. *Ann Oncol*. 2006;17:1783–91.
 9. Breast Screening. Cancer Research UK 2002. Cancer Research UK Charity Number 1089464. Available from: <http://www.cancerhelp.org.uk>.
 10. Shapiro S, Coleman EA, Broeders M, Codd M, de Koning H, Fracheboud J, et al. Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening. *Int J Epidemiol*. 1998;27(5):735–42.
 11. Hyndman RJ, Ullah MS. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput Stat Data Anal*. 2007;51(10):4942–56.
 12. Booth H, Hyndman RJ, Tickle L, De Jong P. Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demogr Res*. 2006;15:289–310.
 13. World Health Organisation Database. WHO 2009. Available from <http://www.who.int>.
 14. Ramsay J, Silverman BW. *Functional Data Analysis*. Springer. 2005.
 15. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press 2003.
 16. Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J Am Stat Assoc*. 1988;83:596–610.
 17. Ramsay JO, Dalzell CJ. Some tools for functional data analysis. *J R Stat Soc, B*. 1991;53:539–72.
 18. Hyndman RJ, Koehler AB, Ord JK, and Snyder RD. Forecasting with exponential smoothing: a state space approach. Berlin: Springer 2008.
 19. Erbas B, Hyndman RJ, Gertig DM. Forecasting age-specific breast cancer mortality using functional data models. *Stat Med*. 2007;26:458–70.
 20. Cuzick J, Powles T, Veronesi U, Forbes J, Edwards R, Ashley S, et al. Overview of the main outcomes in breast cancer prevention trials. *Lancet*. 2003 Jan 25;361(9354):296–300.
 21. Boyle P, d’Onofrio A, Maisonneuve P, Severi G, Robertson C, Tubiana M, et al. Measuring progress against cancer in Europe: has the 15% decline targeted for 2000 come about? *Ann Oncol*. 2003;14:1312–25.
 22. Beral V, Hermon C, Reeves G, Peto R. Sudden fall in breast cancer death in England and Wales. *Lancet*. 1995;345:8965.
 23. Dickinson HO. Cancer trends in England and Wales. Good data and analysis are vital to improving survival. *BMJ*. 2000;320:884–5.
 24. NHS Breast Screening Programme. NHS Cancer Screening Programmes, Fulwood House, Old Fulwood Road, SHEFFIELD S10 Available from: <http://www.cancerscreening.nhs.uk/breastscreen>.
 25. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women’s Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–33.
 26. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clark L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*. 2005;353:1784–92 M.
 27. Møller B, Weedon-Fekjaer H, Haldorsen T. Empirical evaluation of prediction intervals for cancer incidence. *BMC Med Res Methodol*. 2005;5:21.
 28. Wingo PA, Cardinez CJ, Landis SH, Greenlee RT, Ries LA, Anderson RN, et al. Long-term trends in cancer mortality in the United States, 1930–1998. *Cancer*. 2003;97:3133–275.
 29. Richardson DB. The impact on relative risk estimates of inconsistencies between ICD-9 and ICD-10. *Occup Environ Med*. 2006;63:734–40.
 30. Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: comparison of observed with predicted mortality. *BMJ*. 2000;321:665–9.