Journal of Epidemiology
Japan Epidemiological Association

Using functional data analysis models to estimate future time trends of age-specific breast cancer mortality for the United States and England-Wales

Journal:	*Journal of Epidemiology*
Manuscript ID:	JE-2009-0072
Manuscript Type:	Statistical Data
Date Submitted by the Author:	10-May-2009
Complete List of Authors:	erbas, bircan Muhammad, Akram Gertig, Dorota English, Dallas hopper, john kavanagh, anne hyndman, rob
Specialty Area:	Cancer, Theory and Statistics
Keywords:	functional data anlysis models, breast cancer, forecasting

STATISTICAL DATA

Using functional data analysis models to estimate future time trends of age-specific breast cancer mortality for the United States and England-Wales

Bircan Erbas[1], Muhammed Akram[2], Dorota M Gertig[3], Dallas English[4, 5], John L. Hopper[5], Anne M Kavanagh[6] and Rob Hyndman[2]

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

[2] Forecasting Unit, Monash University,
Clayton, 3800, Australia.

[3]Victoria Cytology Service Inc,
Carlton, 3053 Australia.

[4] Cancer Epidemiology Centre
The Cancer Council Victoria
Carlton 3053 Australia

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

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

Corresponding Author: Dr Bircan Erbas
School of Public Health
La Trobe University
Rm 129, Health Sciences 1,
Bundoora, Victoria, Australia 3086
Tel: +61 3 9479 5657
Fax: +61 3 9479 1783
Email: b.erbas@latrobe.edu.au

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

ABSTRACT

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

Material and methods: We use functional data analysis techniques to model breast cancer mortality-age relationships in the United States from 1950 to 2001 and England-Wales from 1950 to 2003, and estimate 20-year predictions using a new forecasting method.

Results: In the United States, trends for women aged 45–54 years continued to decline since 1980. In contrast, trends in women aged 60 - 84 years increased in the 1980s and declined in the 1990s. For England-Wales, trends for women aged 45 to 74 years slightly increased prior to 1980, but declined thereafter. The greatest age-related changes for both countries 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 - 70 years. Forecasts suggest relatively stable trends for women over 75 years.

Conclusions: Predicting 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 modelling and projecting future trends of diseases where there has been very little advancement in treatment and minimal cohort effects such as lethal cancers.

INTRODUCTION

Breast cancer is the most common cancer diagnosed in women worldwide [1]. In the United States and in the UK, 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 resource allocation and policy. In particular, accurate age-specific projections of mortality from breast cancer are essential for assessing the planning of cancer control programs such as mammographic screening. Screening, combined with improvements in treatment options influence 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 compared with women aged 50 to 54 years [6]. Conversely, in the United States breast cancer mortality for white women has declined more rapidly for those under 50 years of age than for older women [5]. A decrease in mortality for women aged 30 to 49 years of age has been observed for a number of European countries without an organized nationwide screening program [4]. Mortality patterns for women over 65 years of age have continued to increase in many countries, regardless of screening and advances in treatment [7, 8].

Differences in age-related mortality trends between countries may reflect differences in mammographic screening policies (i.e. 3-yearly screening of 50 to 70 year-olds in the UK [9] and US has no organised screening program and requires referral by medical practitioner [10] and age-related differences in the uptake of treatment options (such as tamoxifen). Other's factors that may contribute to these differences are HT use and OC use. Given these apparent age related differences, it may be misleading to estimate future breast cancer mortality patterns without taking age-related time trends in mortality into account.

In this study, we use a recently developed forecasting method [11] to: i) compare the time trends of age-specific breast cancer mortality for the United States and England-

Wales; and ii) predict future predictions of age-specific breast cancer mortality for the United States and England-Wales using the age-specific time trend data. The forecasting method we use predicts the entire age-mortality relationship through time rather than using most recent data i.e., a limited time frame to estimate future rates. In doing so, the models are likely to increase the accuracy of these predictions. 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 to 2003 were obtained from the Office of the National Statistics [3]. U.S.A. mortality data from 1950 to 2001 were extracted from the World Health Organization Mortality database [12]. Data for breast cancer are designated by ICD 6 & 7 170, ICD 8 & 9 174, ICD 10 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.

Statistical Methods

Breast cancer mortality is observed annually as a function of age, defined as the midpoint of the age groups. For each year, we plot the age and mortality associations and refer to them mortality-age curves. We take the log of the mortality rate (our outcome) at each mid-point of age for each year and use functional data analysis (FDA) techniques [13] to model these annual mortality-age curves collectively as a functional time series. In these models we first assume an underlying smooth function that we are observing with error (See Appendix for details). We use nonparametric regression techniques to estimate the smooth curves [14, 15].

We then take these smooth curves as our functional observations and fit functional data analysis models. We follow the estimation procedure of Hyndman and Ullah [11] and apply functional principal components decomposition [16] to the smooth curves as this approach gives a small number of basis functions, enables informative

interpretations, and gives coefficients which are uncorrelated with each other. To assess the goodness of fit, we choose a model with basis functions that minimize the mean integrated squared error (MISE) [17].

To predict future mortality, we forecast each coefficient in the models using a univariate time series model. We multiply these forecasts with he basis functions in the FDA models, resulting in forecasts of mortality age curves through time. We use exponential smoothing state space models to compute the forecast [18] and construct prediction intervals around our predictions [17]. We use the Mean Integrated Squared Forecasting Error (See Appendix for details) to evaluate the accuracy of the estimated predictions of future mortalities. For the US and England-Wales mortality data we estimated twenty-year predictions using exponential smoothing state space models with damping [18]. All statistical analyses were performed in 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 1959-2001 (left) and for England-Wales (per 100,000 women) from 1950 - 2003 (right). For the United States, mortality trends for young women (45–54) 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 between the ages 75-79 and 80-84 fluctuated during the study period. However, as is similar to the other age groups, mortality trends declined overall during the 1990s.

INSERT FIGURE 1 HERE

Under our model, an adequate fit (as determined by the MISE) was a functional regression model with two basis functions for the United States, while for England-Wales it was a functional regression model with three 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 the England –Wales. For England-Wales, β_1 shows an increase in the mortality at all age groups between 1950 to 1980, followed by a rapid decline until 2000. Similarly interpretations can be made for the USA.

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: a) the United States from 2002 to 2021; and b) England-Wales from 2004 to 2023. The y-axis represents the coefficients associated with the first basis function i.e., this controls the overall change in trend of breast cancer mortality. Predictions from the models suggest that overall crude mortality rates for both countries are expected to decline more slowly than was observed during the 1990s, and stabilize around or beyond 2020. These predictions assume no changes/advances in treatment.

INSERT FIGURE 2 HERE

Figure 3a and 3b displays twenty-year predictions of age-specific breast cancer mortality for both countries. Mortality trends are expected to decline for all women, with the greatest decline for women 60 to 70 years of age, whereas estimated predictions suggest relatively stable trends for women over 75 years

INSERT FIGURE 3a AND 3b HERE

To evaluate the accuracy of the predictions, we estimated 1-, 10- and 20-year age-specific mortality predictions with 80% prediction intervals for both countries (Figure 4). The estimated predictions had remarkably narrow prediction bands; for example, twenty-year predictions for 60 year old women in England-Wales had an 80% error margin of less than 10 deaths per 100,000 women.

INSERT FIGURE 4 HERE

We examined the residual of the functional fits using image plots (data not shown). These image showed no evidence of lack of fit and suggest very little remaining birth cohort trends to be of no concern.

DISCUSSION

Using an innovative forecasting method the twenty-year projections suggest a continuing decline which will stabilise in overall mortality for both England-Wales and the United States. Major factors contributing towards the continuing decline in mortality observed for England-Wales, and to a lesser extent for the United States, have been widely debated [19, 20]. 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 [19]. Screening has also been considered to have played a role towards the end of the 1990s and more recently, but there are doubts as to whether it had a major impact on mortality in the early 1990s when the sharpest

decline occurred [21, 22]. The UK has had an organised mammographic screening program targeting women aged 50 to 70 years in place since 1988 [23] whereas in the United States screening is more ad-hoc and recommendations are for women over 40 years of age [20]. It is possible that screening contributes to the greater decline in women >75yrs in UK but not US. Other factors, such as the rapid increase in hormone therapy use in the 1990s and the subsequent falls following the release of the Women's Health Initiative trial findings [24] may have contributed to the continuing decline in breast cancer mortality for England-Wales. The observed reduction for 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 [25].

This study has a number of important strengths. First we have presented an alternative modelling 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 modelling and forecasting approach is appealing because of the few parameter assumptions that need to be made and the visual interpretation of the projections. In contrast to the common age-period-cohort model, our models represent the age related changes in mortality as response curves. In APC models an overall effect for age is estimated so that variations in mortality rates in younger women with those of middle aged and older women are combined and thereby underestimating age-related changes in mortality. As a result, estimated mortality predictions may be inaccurate.

Second, unlike most other studies, we estimate prediction intervals for future mortality rates. Prediction intervals are necessary to accommodate both the uncertainty in the mathematical structure of the model and the variation in the future mortality rate [26]. The prediction intervals from the functional forecasting models are narrow for both the overall and the 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, we observe wider prediction intervals as the forecast horizon increases.

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 through the years, revisions of ICD codes, and the combination of subsites into one major site may all affect our interpretation of cancer trends [27] however for breast cancer there is general consensus that the consistency has been reasonable [28].

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 [26]. Birth cohort trends for US women born between 1880 through to 1915 would have impacted on the estimates and thereby making projections higher than expected for older women. At present, the decreasing birth cohort trends for baby boomers have not made a full impact on the decline in future mortality rates in the US. In a study of the mortality benefits of screening in England-Wales [29] birth cohort effects were similar for all ages prior to the introduction of screening. The smoothing process as part of the modelling 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 modelling mortality/incidence trends and we are currently developing the models to incorporate other variables that may influence age-related changes in mortality (or incidence).

In summary, we present a new modelling and forecasting technique to model and estimate future trends in breast cancer mortality. At present, much of the modelling and prediction of mortality trends focus on ACP models. Here, we present an alternative framework with features such as predicting the entire age-mortality curves for each period to estimate the predictions, thereby enhancing the accuracy of the predictions. In their current form, these models will be most useful for modelling and projecting future trends of diseases where there has been very little advancement in treatment and minimal birth cohort effects such as pancreatic or brain cancers.

REFERENCES

1. Parkin DM, Pisani P, Ferlay J. Global Cancer statistics. CA Cancer J Clin 1999; 49: 33-64
2. http://www.cancer.org/statistics
3. http://www.statistics.gov.uk
4. Tycynski JE, Plesko I, Aareleid T 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: 1056-1064
5. Smigal C, Jemal A, Ward E et al. Trends in breast cancer by race and ethnicity: update 2006. CA Cancer J Clin 2006; 56:168-83
6. Jatoi I and 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 et al. 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. www.cancerhelp.org.uk
10. Shapiro S, Coleman EA, Broeders M 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; 27:735-42.
11. Hyndman RJ and Ullah MS. Robust forecasting of mortality and fertility rates: a functional data approach. 2007. Computational Statistics and Data Analysis. Accepted 5July 2006.
12. www.who.int
13. Ramsay J and Silveramn BW. Functional Data Analysis. Springer. 2005
14. Ruppert D, Wand MP, and Carroll RJ. Semiparametric Regression. Cambridge University Press 2003.
15. Cleveland WS and Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J Amer Stat Assoc 1988; 83: 596-610
16. Ramsay JO and Dalzell CJ. Some tools for functional data analysis. J Roy Stat Soc. Series 1991; 53: 539-572

17. Erbas, B, Hyndman RJ and Gertig DM. Forecasting age-speci.c breast cancer mortality using functional data models. Stats in Med 2007; 26: 458–470.
18. Hyndman RJ, Koehler AB, Synder RD et al. A state space framework for automatic forecasting using exponential smoothing methods. Int J Forec 2002; 18: 439-454
19. Brown P. UK death rates from breast cancer fall by a third. BMJ 2000; 321: 849
20. Boyle P, d'Onofrio A, Maisonneuve P et al. Measuring progress against cancer in Europe: has the 15% decline targeted for 2000 come about? Ann Oncol 2003;14:1312-25
21. Beral V, Hermon C, Reeves G et al. Sudden fall in breast cancer death in England and Wales. The Lancet 1995; 345: 8965
22. Dickinson HO. Cancer trends in England and Wales. Good data and analysis are vital to improving survival. BMJ 2000;320:884-5
23. www.cancerscreening.nhs.uk/breastscreen
24. Rossouw JE, Anderson GL, Prentice RL 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.
25. Berry DA, Cronin KA, Plevritis SK et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005; 353:1784-92. M
26. Møller B, Weedon-Fekjaer H, Haldorsen T. Empirical evaluation of prediction intervals for cancer incidence. BMC Med Res Methodol 2005;5:21.
27. Wingo PA, Cardinez CJ, Landis SH et al. Long-term trends in cancer mortality in the United States, 1930-1998. Cancer 2003;97:3133-275
28. Richardson DB. The impact on relative risk estimates of inconsistencies between ICD-9 and ICD-10. Occup Environ Med 2006; 63:734-40.
29. Blanks RG, Moss SM, McGahan CE et al. 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

APPENDIX A

We use functional data analysis techniques Ramsay and Silverman [14] to model these 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 $\{x_i, y_t(x_i)\}$, $t = 1,\ldots,n$, $i = 1,\ldots,p$, where

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

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 splines Ruppert, Wand and Carroll [15] and loess curves Cleveland and Devlin [16]. The n smoothed curves are our functional observations, $\{f_t(x)\}$, where $x_1 < x < x_p$ and $t = 1,\ldots,n$. Erbas, Hyndman and Gertig [18] and Hyndman and Ullah [12] proposed the following model for the smoothed curves:

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

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 the functional principal components decomposition Ramsay and Dalzell [17] applied to the smoothed curves $\{f_t(x)\}$, as this approach gives a small number of basis functions, enables informative interpretations, and gives coefficients which 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 mortalities, 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 $f_{n+h}(x)$, $h=1,2,\ldots$.We use exponential smoothing state space models to compute the forecast (see Hyndman et al. [19]), and construct prediction intervals around our predictions (see Erbas, Hyndman and Gertig [18], for a detailed description). The Mean Integrated Squared Forecasting Error: (MISFE(h)) $=\frac{1}{n-m+1}\sum_{t=m}^{n}\int\left[y_{t+h}(x)-\hat{f}_{t,h}(x)\right]^2 dx$ is used to evaluate the accuracy of the estimated predictions of future mortalities.

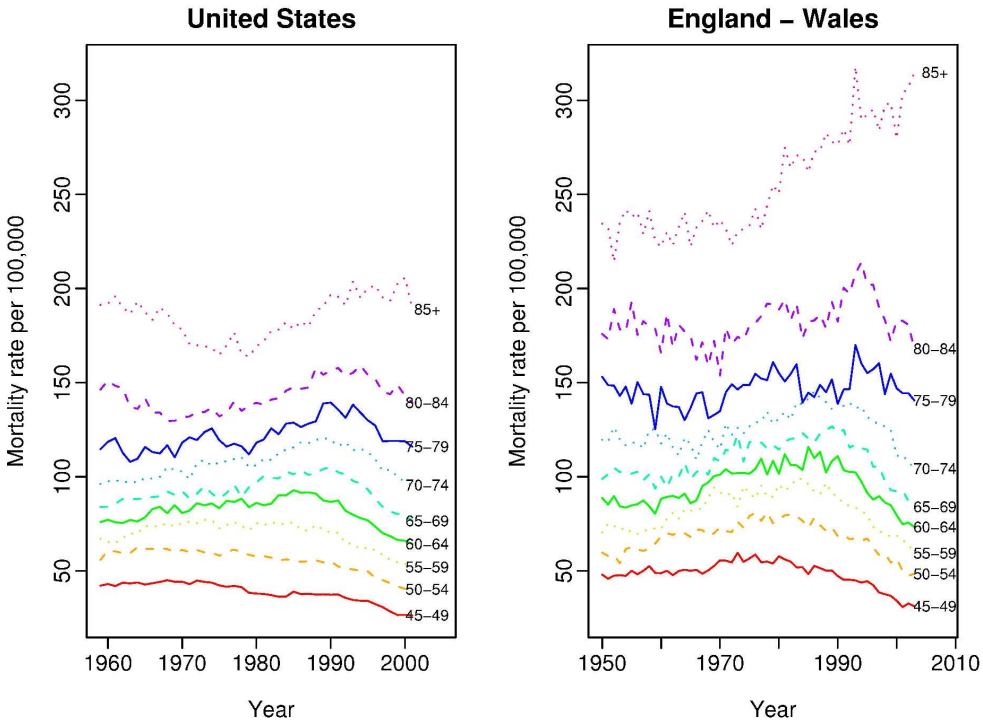
Figure 1:Observed breast cancer mortality trends by age group for the United States (left) and England-Wales (right)

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.

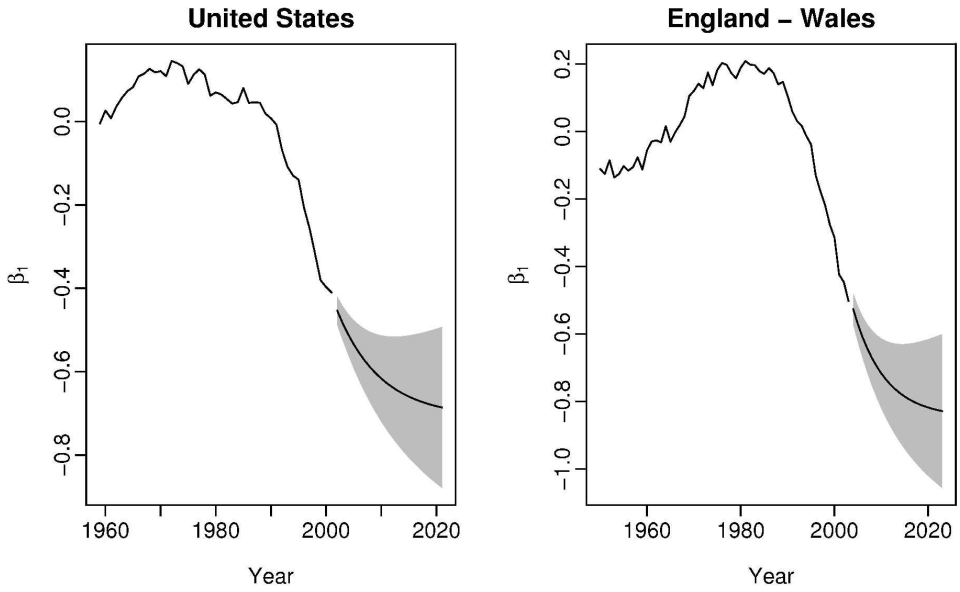
Figure 3a: Estimated twenty-year predictions of age specific breast cancer mortality for the United States

Figure 3b: Estimated twenty-year predictions of age specific breast cancer mortality for England and Wales

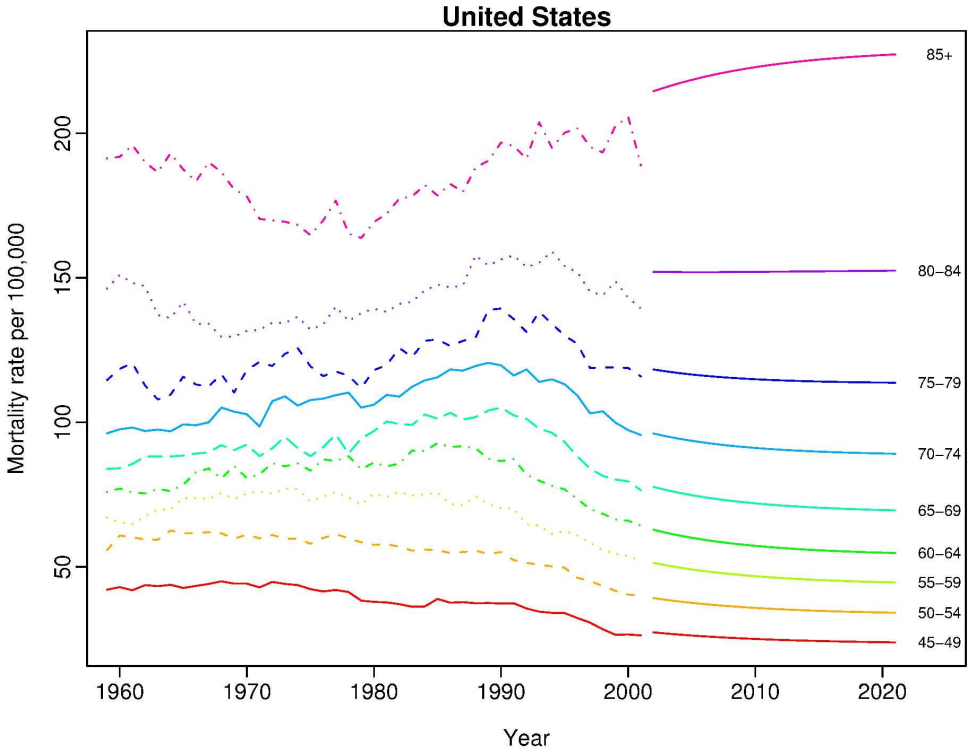
Figure 4: Estimated 1, 10 and 20 year age specific predictions for the United States (top) and England-Wales (bottom)



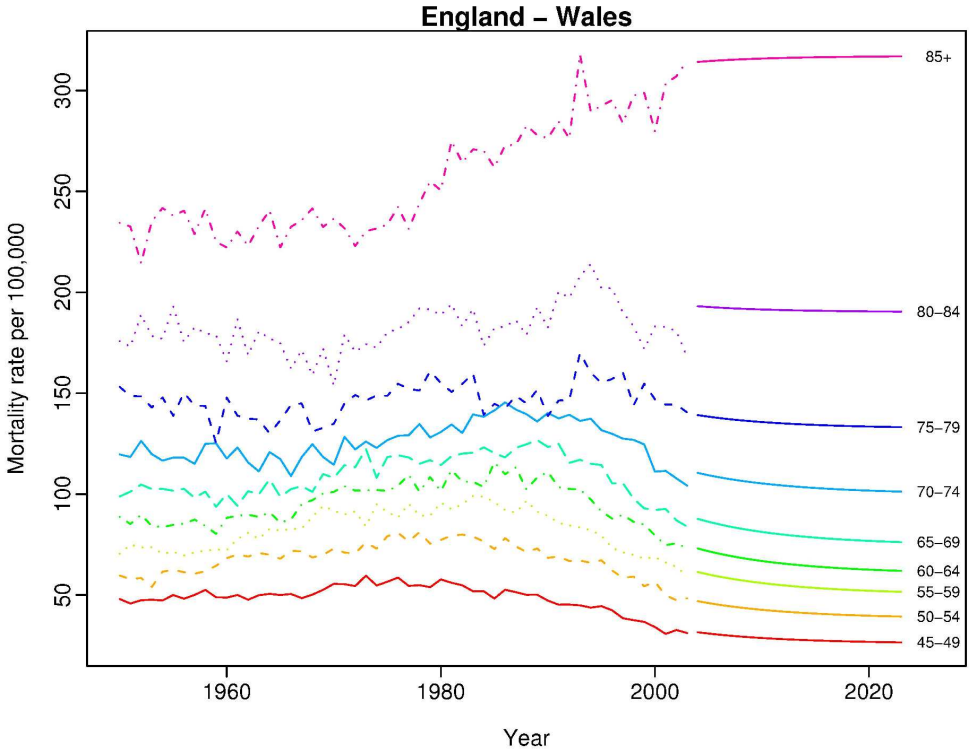
159x119mm (600 x 600 DPI)



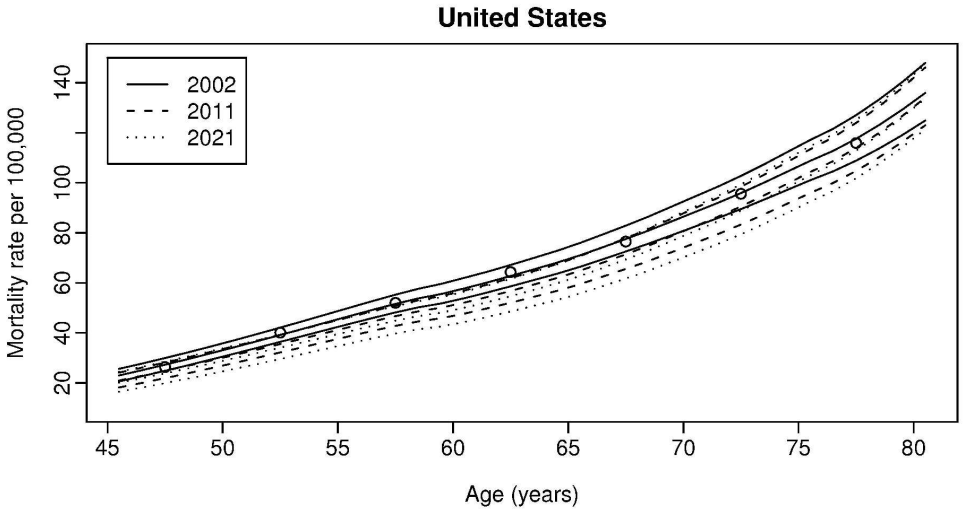
161x100mm (600 x 600 DPI)

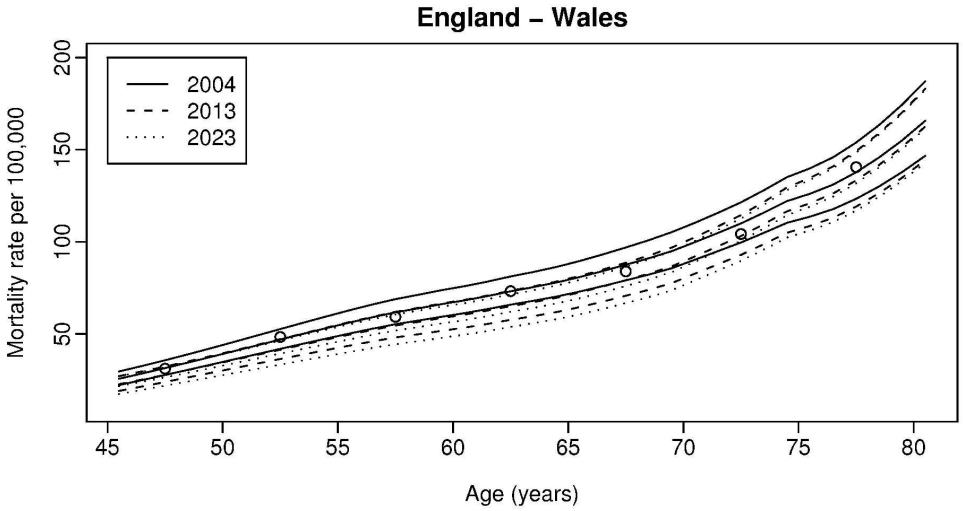


159x119mm (600 x 600 DPI)



159x119mm (600 x 600 DPI)





159x179mm (600 x 600 DPI)