FORECASTING MORTALITY, DIFFERENT APPROACHES FOR DIFFERENT CAUSE OF DEATHS? THE CASES OF LUNG CANCER; INFLUENZA, PNEUMONIA, AND BRONCHITIS; AND MOTOR VEHICLE ACCIDENTS

By Mariachiara Di Cesare and Mike Murphy

ABSTRACT

Most of the methods of mortality forecasting have been assessed using performance on overall mortality, and few studies address the issue of identifying the appropriate forecasting models for specific causes of deaths. This study analyses trends and forecasts mortality rates for three major causes of death — lung cancer, influenza-pneumonia-bronchitis, and motor vehicle accidents — using Lee–Carter, Booth–Maindonald–Smith, Age-Period-Cohort, and Bayesian models, to assess how far different causes of death need different forecasting methods. Using data from the Twentieth and Twenty-First Century Mortality databases for England and Wales, results show major differences among the different forecasting techniques. In particular, when linearity is the main driver of past trends, Lee–Carter-based approaches are preferred due to their straightforward assumptions and limited need for subjective judgment. When a clear cohort pattern is detectable, such as with lung cancer, the Age-Period-Cohort model shows the best outcome. When complete and reliable historical trends are available the Bayesian model does not produce better results than the other models.

CONTACT ADDRESS

Mariachiara Di Cesare, Department of Social Policy, London School of Economics, Houghton Street, London WC2A 2AE, U.K. Tel: +44-20-7955-7698;
E-mail: mariachiara.dicesare@gmail.com

1. INTRODUCTION

Cause of death forecasts are increasingly important due to the implications of changing cause of death patterns for health and social care costs predictions as well as for their contribution to understanding the drivers of overall mortality change (Crimmins, 1981; Scitovsky, 1994; Tabeau *et al.*, 1999; Polder *et al.*, 2006).

In recent decades a range of methods for mortality forecasting has been developed (e.g. Lee & Carter, 1992; Heathcote & Higgins, 2001; Booth *et al.*,

2002; Renshaw & Haberman, 2006) including techniques for the estimation of uncertainty of forecasts (Lutz & Goldstein, 2004; Girosi & King, 2008); for a review see Booth & Tickle (2008). Most applications and comparison of models have been based on performance using overall mortality (Booth *et al.*, 2006), and few studies address the issue of establishing the appropriate forecasting models for specific causes of deaths.

This study analyses trends and forecasts mortality rates for three major causes of death that have different underlying trends and drivers: malignant neoplasm of trachea, bronchus and lung ('lung cancer' subsequently); influenza, pneumonia and bronchitis ('IPB' subsequently); and motor vehicle accidents ('MVA' subsequently). These three causes were selected since the relative importance of cohort and period factors and year-to-year changes vary between them. While lung cancer has a strong cohort influence, influenza, pneumonia and bronchitis is mainly characterised by period factors, while motor vehicle accidents does not have either clear cohort or period patterns (see Figures A1 and A2 in Appendix). Lung cancer shows the smoothest pattern and IPB the largest fluctuations with time.

Three different models (plus a variant), one from each of different families of forecasting techniques (Lee–Carter, Booth–Maindonald–Smith, Age-Period-Cohort, and Bayesian models) are applied to assess how far different causes of death need different forecasting methods, and the extent to which alternative methods provide insights into the underlying processes. Indicators of model goodness of fit and of forecasting performance are used to assess the best model for each selected cause of death.

2. The Models

2.1 The Lee–Carter Model

The Lee–Carter (LC) method (Lee & Carter, 1992) is the most widely-used method for forecasting future mortality, used for example by the U.S. Census Bureau for official U.S. population projections (Hollmann *et al.*, 2000). Its strengths are simplicity and robustness especially in the case of largely linear time trends in age-specific death rates (Booth *et al.*, 2006). The method consists of a base model of age-specific death rates with a fixed relative age component and a dominant time component and forecast using time series approaches (Booth *et al.*, 2002).

The Lee–Carter model for each sex and for a particular cause of death is:

$$\ln(m_{at}) = \alpha_a + \beta_a k_t + \varepsilon_{at} \tag{1}$$

where
  m_{at} is the death rate at age a at time t,
  k_t is an index of the overall level of mortality at time t,

α_a are the age-specific mean values of $\ln(m_{at})$ averaged across the years over which the model is fitted,

β_a is the relative speed of change at each age, and

ε_{at} is the residual at age a and time t.

β_a and k_t are estimated by singular value decomposition. To obtain a unique solution for the model, constraints need to be imposed: conventionally the estimated β_a values sum to 1, and the estimated k_t values sum to zero. The Lee–Carter model incorporates an adjustment to the estimates of k_t so that the sum of fitted deaths matches the observed total deaths in each year; this gives greater weight to ages at which deaths are high thereby partly counterbalancing modelling on the logarithmic scale that also gives considerable weight to ages with low mortality rates. The time series of k_t values is forecast using a standard ARIMA time series forecasting approach, although in practice usually as a simple random walk with drift model is found adequate (Booth *et al*., 2006).

$$k_t = k_{t-1} + d + e_t \qquad (2)$$

where

d is the average annual change (drift) in the k_t series, and

e_t values form a set of uncorrelated terms.

The Lee–Carter model is widely adopted because of its advantages such as: no subjective judgment involved; forecasting based on long term trends; and availability of confidence intervals (Lee & Carter, 1992). However, a major problem with the LC model is the assumption that the age pattern is invariant over time and hence no age-time interaction is taken into account (Lee & Miller, 2001).

2.2 *The Booth–Maindonald–Smith Variant of the Lee–Carter Model*

The Booth–Maindonald–Smith (BMS) is a variant of the Lee–Carter approach that differs in three main aspects. "The fitting period is chosen based on statistical goodness-of fit criteria under the assumption of linear k_t; the adjustment of k_t involves fitting to the age distribution of deaths; and the jump-off rates (the rates in the last year of the fitting period) are taken to be the fitted rates based on this fitting methodology" (Booth *et al*., 2006, p293). The BMS model was developed after the analysis of mortality decline in Australia during the twentieth century. Booth *et al*. (2006) found that the BMS variant forecast Australian mortality more accurately than the classical Lee–Carter method, however they found no significant differences in quality of BMS forecasts over a number of countries compared with other variants such as Lee & Miller (2001), Hyndman & Ullah (2007), or the De Jong & Tickle (2006) method. The Lee & Miller (2001) application restricts the

fitting period in the U.S. from 1950 to be more consistent with the model assumption, adjusts k_t based on life expectancy at birth in year t, and considers the jump-off rates as the actual rates in the jump-off year. The Hyndman & Ullah (2007) application extends the Lee–Carter method by assuming mortality is a smooth function of age (nonparametric methods are used), by considering more than one set of (k_t, β_a) values, using more general time series methods, and incorporating robust estimation to control for 'exceptional' years, and not adjusting k_t. The De Jong–Tickle method uses state space models of which Lee–Carter is a special case.

2.3   *The Age-Period-Cohort Model*

Age-Period-Cohort models (APC) have been developed to estimate age, period, and cohort patterns in time series, including a number of mortality studies, and they continue to be used for this purpose (Mason & Smith, 1985; Yang *et al.*, 2008; Cleries *et al.*, 2009). The general APC model for the log-death rates, m_{atc}, at age a in time period t for persons in birth cohort $c = t - a$, is:

$$\ln(m_{atc}) = f(a) + g(t) + h(c). \tag{3}$$

The age, period and cohort functions can be simple dummy variables or a function of the corresponding variables. In recent applications, these are often modelled as smooth spline functions (e.g. Carstensen, 2007). One of the main limitations of APC models is the identification problem: each effect — age, period, and cohort — is linearly linked to the other two remaining effects since $c = t - a$ (so that if any of the two of the a, t or c components are known, the third one is known exactly).

According to Tabeau (2001, p16): "an APC model can be classified as symptomatic: period effects approximate contemporary factors, such as health status of the population, and cohort effects approximate historical factors". The APC model is best considered as a descriptive tool (Carstensen, 2007; Booth & Tickle, 2008) of past patterns with specific limitation as a forecasting method in that even if age and cohort effects can be assumed to be fixed, it is usually not feasible to assume fixed period effects (Tabeau, 2001). The model is usually fitted as a Poisson or related family such as the negative binomial model for death counts, with exposure as an offset within a Generalised Linear Model (GLM) framework.

2.4   *The Bayesian Model*

The Bayesian approach to mortality forecasting allows incorporation of levels of uncertainty in the model. Girosi & King (2008) developed a Bayesian hierarchical model able to pool information from similar cross-sections such as age groups or time in an efficient way. Under the Bayesian approach it is possible to incorporate *a priori* information about the data.

This prior information is then used together with the observed data to define the posterior distribution on which inference is based (Pedroza, 2006). The prior knowledge used in this case is that the expected value of the dependent variable μ_{at} — where it is assumed that the observed value $m_{at} \sim N(\mu_{at}, \sigma^2)$ — varies smoothly over age, 0 to A, time, 0 to T, and age/time according to the following prior:

$$
\begin{aligned}
H\big[\mu, \theta_{age}, \theta_{time}, \theta_{age/time}\big] &\equiv \frac{\theta_{age}}{AT} \int\limits_0^T dt \int\limits_0^A da \left(\frac{d^2}{da^2} (\mu(a, t) - \overline{\mu}(a)) \right) \\
&+ \frac{\theta_{time}}{AT} \int\limits_0^T dt \int\limits_0^A da \left(\frac{d^2}{dt^2} \mu(a, t) \right) \\
&+ \frac{\theta_{age/time}}{AT} \int\limits_0^T dt \int\limits_0^A da \left(\frac{\partial^3 \mu(a, t)}{\partial a \partial t^2} \right). \quad (4)
\end{aligned}
$$

The first functional smoothes over age groups; the second smoothes the time series over time ensuring that the curvature of the time series stays within reasonable bounds (based on integral of the second derivative as commonly used in smoothing algorithms). The last term is a mixed age-time smoothness functional ensuring that the curvature of the time series varies smoothly over age groups (Girosi & King, 2008). The Bayesian hierarchical model works better when applicable *a priori* knowledge is incorporated in the model. In practice, no major benefit is added to the forecast if no prior quantitative or qualitative analysis or knowledge exists.

3. Data

The data used for the analysis comes from the ONS Twentieth Century Mortality database for England and Wales, updated with the 21st Century Mortality files (Office for National Statistics, 2000 and 2009). Central mortality deaths rates were computed for the three causes considered: malignant neoplasm of trachea, bronchus and lung (International Classification of Diseases Revision 10 (ICD10) codes: C33-C34 (or equivalent codes in earlier version of ICD); influenza, pneumonia, and bronchitis (ICD10 codes: J10-J22); and motor vehicle accidents (ICD10 codes: V02-V04, V09.0, V09.2, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2), and mid-year population by sex and five-year age classification (age groups 0-4 split into age zero and 1-4 and open-ended

band ages 85 and over). To avoid problems with the linearity assumption implicit in the LC models and ensure the highest level of comparability in consistency of death coding among the different models, the methods have been fitted to the period 1950-2007, and forecast for the thirty years forward, 2008-2037, both for the cause-specific and overall cases, since forecasts at longer horizons are increasingly unreliable. All analyses were undertaken using the R statistical package (R Development Core Team, 2009). Some restrictions on ages included have been applied according to the cause of mortality analysed: for lung cancer, the models start from age 35. The results show the sorts of age-specific patterns that are obtained with the various models using examples of causes of death trends with differing underlying processes. One basis for assessing forecasts is the extent to which the patterns produced are judged to be plausible in the light of previous experience and current knowledge of how these may evolve.

To assess the models' forecasting performance using an alternative approach, models have also been fitted over the period 1950-1977 which permits out-of-sample forecasts to be compared with actual values for the 30-year period 1978 to 2007. The mean absolute errors (MAE) of the observed and forecast values of the logarithm of the age-specific mortality rates by age, time and overall value are used as the goodness-of-fit measure in the period 1978-2007. Since differences of the logarithm reflect proportional rather than absolute differences in mortality rates, the same proportionate error at high and low rates will receive equal weight in the calculations, but may have very different impacts on other — possibly substantively more important — measures such as life expectancy. Therefore comparison between actual and forecast period life expectancy at birth based on overall mortality in this period will also be used as an additional indicator to assess the performance of each method.

4. CHANGES IN THE MORTALITY PROFILE DURING THE PAST CENTURY

Figure 1 shows age standardised death rates (SDR, using the European Standard Population) for malignant neoplasm of trachea, bronchus, and lung (Lung), Influenza-Pneumonia-Bronchitis (IPB) and motor vehicle accidents (MVA) plotted for periods when consistent ICD codes for each specific cause are available. The lung cancer historical trend (Figure 1, first panel) for males shows a sharp increase between 1930 and 1970, but starting from the middle of the 1970s the trend decreases at a similar pace as the earlier increase, so that the actual level of the age standardised death rate in 2007 had more than halved from its peak to around 50 per 100,000. Compared with males, the female pattern is shifted in time and intensity, with the peak in the age standardised death rate observed around 1985 at about 30 per 100,000 and it does not show a substantial subsequent decrease, so that the
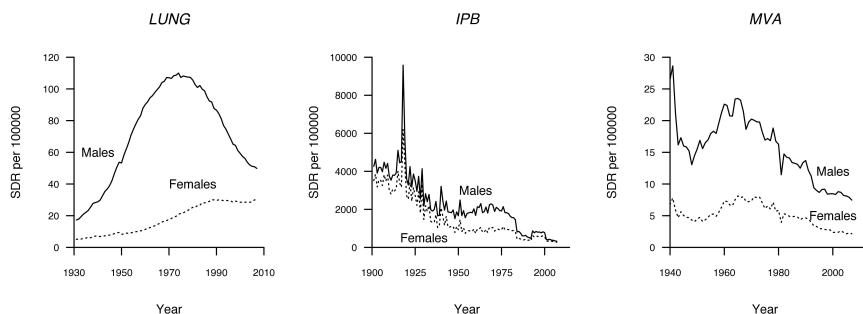
Source: Office for National Statistics (2000, 2009)

Figure 1.   Age standardised death rate for malignant neoplasm of trachea, bronchus, and lung (Lung), Influenza-Pneumonia-Bronchitis (IPB), Motor vehicle accidents (MVA) per 100,000, Males and Females. England and Wales

current level of the age standardised death rate for females is two-thirds of that for males, while at the time of the male peak, the early 1970s, it was just one-fifth of the corresponding value.

The shift in the pattern of lung cancer mortality trends and the difference in levels in recent decades between men and women are mainly due to differences in smoking behaviours. Cigarette use among women started later than among men in Great Britain, with a general lower prevalence (Davy, 2006). The proportion of people smoking cigarettes fell between the early 1970s and 2004/5, and the gender gap has also reduced; in 1972 it was estimated that 52% of men (aged 16 and over) and 41% of women smoked; by 2004/5 the values were 26% for men and 23% for women. The gradual reduction in lung cancer mortality observed among both sexes is due to fewer young people starting to smoke, and smokers giving up at younger ages (Davy, 2006).

The IPB age standardised death rate trend (Figure 1, middle panel) shows a general decrease for both men and women, broken by periodic epidemics, such as the Spanish flu pandemic in 1918-19 and during the Second World War (Griffiths & Brock, 2003). Between 1979 and 2000 an irregular pattern of jumps in the trend of the Influenza-Pneumonia-Bronchitis deaths series is visible, due to a change in coding rules occurred within ICD9 which affected the attribution of leading cause of death for older people. A similar discontinuity occurs in the last seven years with the introduction of ICD10, due to changes in coding rule 3, which allows whatever condition reported in both Part I and II of the death certificate to take precedence over the condition selected following the other rules if it is clearly a direct

consequence of that condition (Brock *et al*., 2006), so the number of deaths due to respiratory diseases decreased by 22% (the effect increasing with age).

The age standardised death rate for motor vehicle accidents (Figure 1, right panel) is shown from 1940, the first year in which the code was included in ICD. Excluding the first 10 years (which correspond to the ICD5 coding period and includes the very different pattern of war-time vehicle use), the trends have an inverted U-shaped profile with an increase in motor vehicle accident deaths during the 1950s and 1960s when car use was becoming more common. The decrease from the 1970s in the death rate corresponds to a time when more strict regulations (such as drink driving) and better safety systems (both roads and in cars, such as seat belts) were introduced. Levels of mortality were lower among females than males, however the gender gap is tending to reduce.

5.   COMPARATIVE FORECASTING PERFORMANCE

To obtain the best fit for each of the four model types included, some assumptions about model parameters are required. In the Bayesian model, non-linear rates of increase in log-mortality have been included (Girosi & King, 2008) for the three causes of death considered. It is particularly important to include such a non-linear term in cases such as motor vehicle accidents, where a linear trend fails to capture the influence of the factors that may initially drive mortality up (or down), but subsequently drive mortality down (or up).

For the Age-Period-Cohort forecasting model, natural splines have been used to model the age, period and cohort functions. The mortality rate in the APC model can be expressed as:

$$m_{atc} = m_a \cdot RR_t \cdot RR_c \qquad (5)$$

where m_a is the age-specific mortality rate for the arbitrary reference cohort, the RR_t is the relative rate referring to the period, t, and RR_c is the relative rate referring to the cohort, c. This model includes 1900 as the reference cohort, which means that the age function can be interpreted as the log-mortality rate in that cohort after adjustment for the period effect. Therefore the cohort function is zero at 1900 (it is the logarithm of the rate ratio relative to that cohort in this formulation), and the period function (for which no reference year has been included) is the logarithm of the rate ratio relative to the age-cohort prediction (Carstensen, 2007).

In addition, for the forecast values, a specific concern arises. The APC model estimates the cohort function as far back as cohorts born before 1865 (1950 minus the age of the oldest age group aged 85 years and over) up to the latest cohort born in 2007 (2007 minus the age of the youngest age group).

The estimation of the cohort function obtained for the youngest cohorts as well as for the oldest cohorts is based on very limited information — for example the cohort estimate for the 2007 cohort is based only on the single death rate at age 0 for people who died in 2007. The estimated values for the youngest cohorts are potentially biased since they are based on a limited number of observations at the tail of the age distribution (the same is true for the oldest cohorts at the other end of the age distribution). The model assumes that values estimated at very young ages for these cohorts will continue to hold at all older ages in years to come. Although some studies have found associations between changes in infant and toddler and later-age mortality in historical cohorts (Catalano & Bruckner, 2006; Murphy, 2010), early-age mortality has not been a good predictor of later adult mortality for recent cohorts (Finch & Crimmins, 2004; Barbi & Vaupel, 2005). For this reason the projection of the cohort function has been based on values for cohorts born up to 1970 (i.e. aged at least 37) by linearly interpolating the cohort function between 1940 and 1970. The adjusted projected cohort functions (except for malignant neoplasm of trachea, bronchus and lung in which the youngest cohort included in the analysis is the one born in 1972) are consistently larger than the initial estimates (see Appendix Figures A1 and A2), implicitly assuming that the recent high rates of cohort mortality improvement estimated at young ages will not be maintained in years to come, and to that extent, they are conservative. This approach will lead to some inconsistency in the forecast rates around the time of the jump-off year among the youngest age groups. Although this choice affects forecasts in the years immediately following the observed period among these groups, long-term use of estimated values based on the very limited experience of recent cohorts can lead to potentially biased results. Overall cohort values will largely reflect patterns at ages where high proportions of deaths occur, but by 2007, 98% of those born 18 years earlier were still alive. This suggests that more weight should be given to results from those cohorts that include more of the main mortality ages even if they are further from the forecast jump-off year than cohorts that include only deaths at the youngest ages. To better appreciate the goodness-of-fit of the APC models, the MAE has been calculated also on the age groups over 35 years (see Table 1).

In Figure 2 (males) and Figure 3 (females), the observed and forecast values for malignant neoplasm of trachea, bronchus and lung are plotted for selected age groups over time. Clearly the historical trend of this cause of death shows a non-linear pattern (an inverted U-shape) especially at ages where rates are high. The male trend increases between 1950 and 1970 and then decreases as sharply from the middle of the 1970s, whereas the female pattern is shifted in time and is lower in intensity, reaching a peak around 1985. In recent decades men are characterised by a decreasing trend (Figure 2), while for women (Figure 3) the observed trend is still increasing in the three oldest age groups.
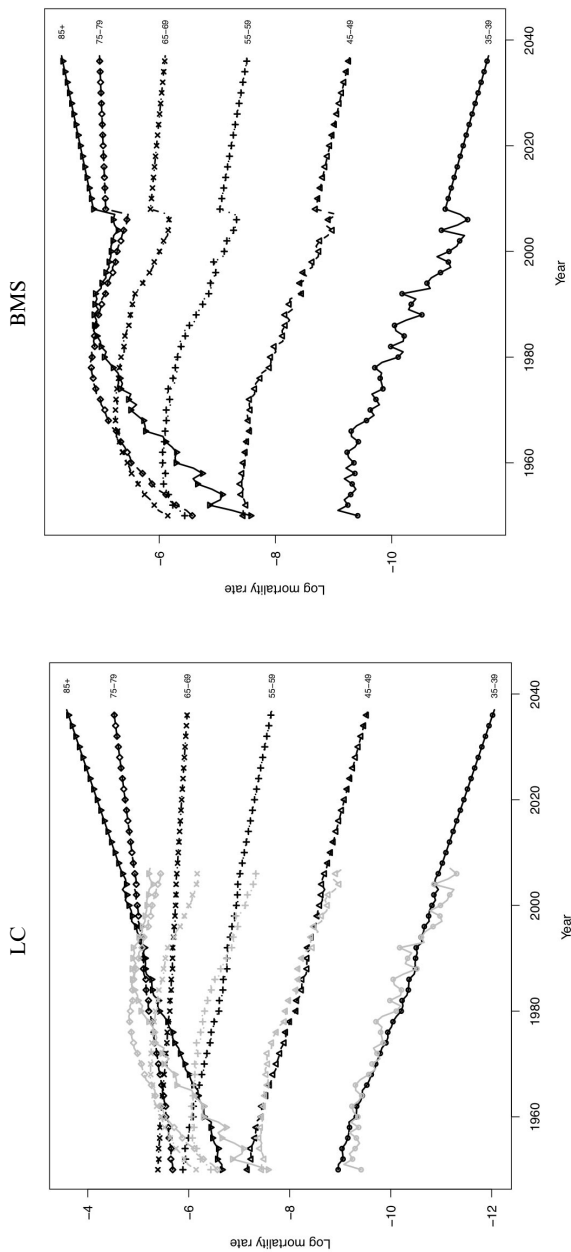
Notes: For clarity, only selected age groups are shown. All models are fitted to years 1950-2007. For BMS model no fitted values are shown due to the jump-off year roles, which would mislead the interpretation of the fitted versus observed values.

Figure 2. Malignant neoplasm of trachea, bronchus, and lung log-death rates by age, observed (light grey), fitted, and forecast values — LC, BMS, Bayesian, and APC model — Males
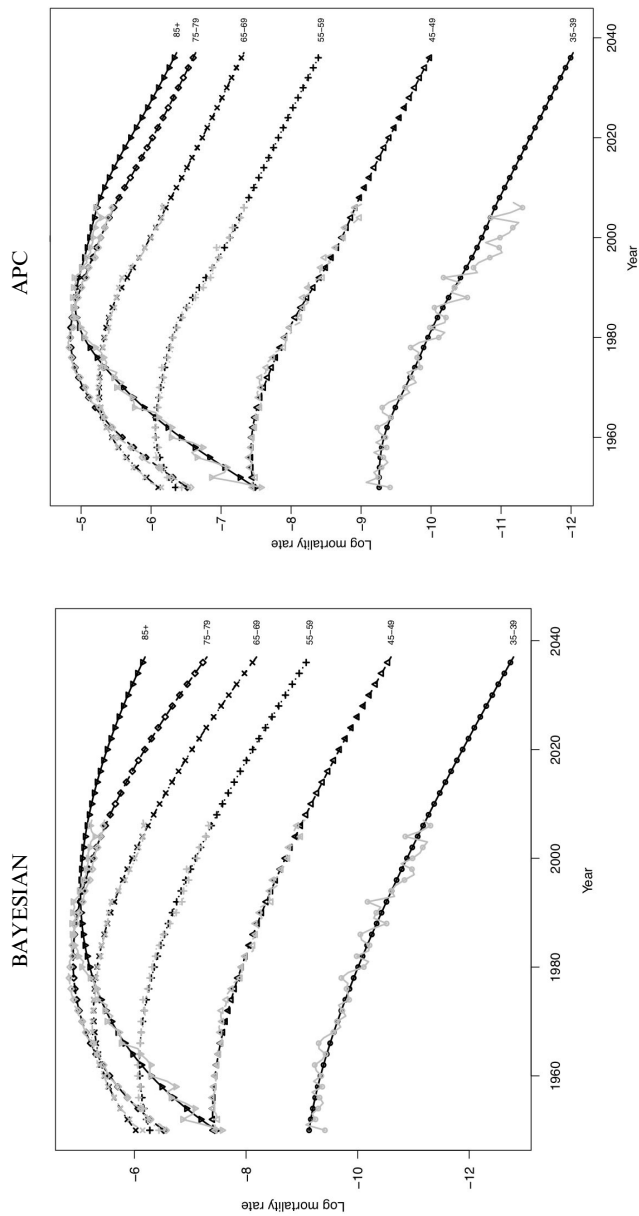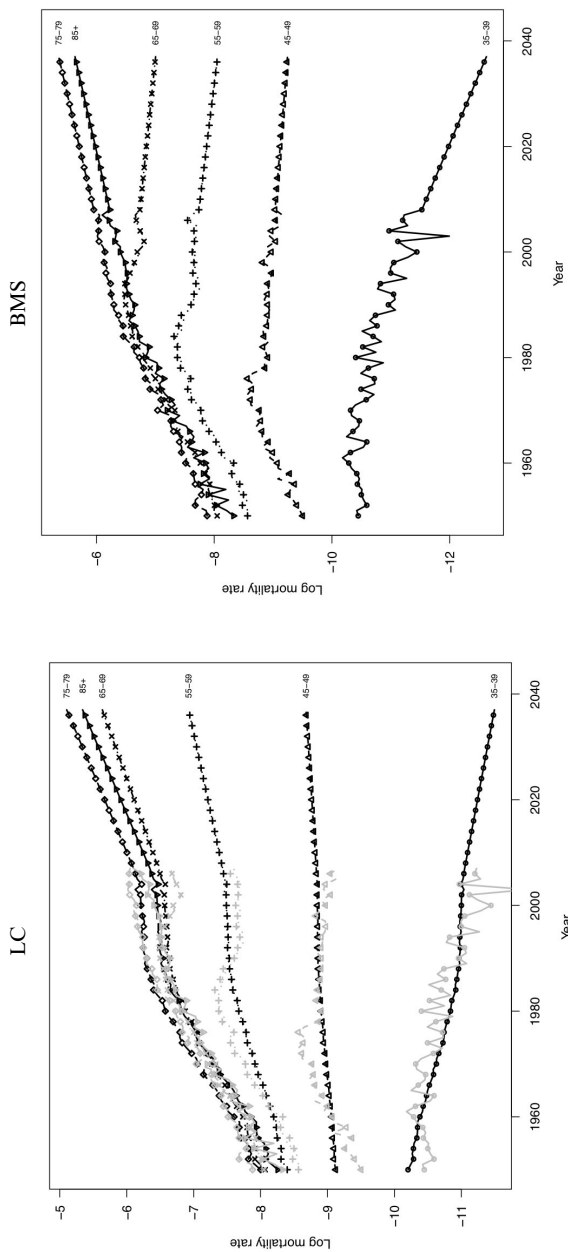
Figure 2. Continued

Note: See Notes to Figure 2.

Figure 3.   Malignant neoplasm of trachea, bronchus, and lung log-death rates by age, observed (light grey), fitted, and forecast values — LC, BMS, Bayesian, and APC model — Females
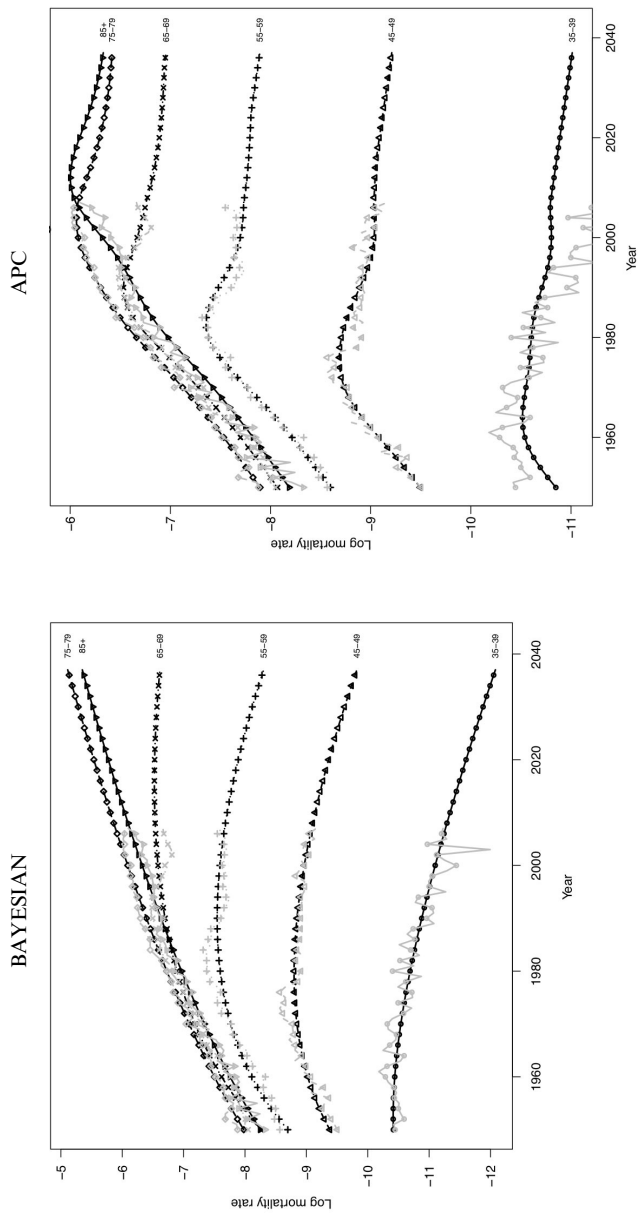
Figure 3.   Continued

Both Figures clearly show that the LC family models allows independence of forecasts for the different age groups while the Bayesian model constrains the forecast age patterns to follow similar paths for adjacent age groups.

The graphs show that both the LC and BMS models are unable to capture the inverted U-shape, since the random walk with drift model is based on an assumption of linearity and consequently both produce inconsistent forecast results. The Bayesian model (without covariates) is able to model the inverted U-shape and to produce a forecast that is visually consistent with the observed trend for men (Bayesian labelled chart in Figure 2), but in the female case where the oldest age groups (70 and over) are characterised by an increasing trend the model fails to forecast the inverted U-shape. Finally, the APC model, which combines the forecast cohort function (which decreases after the peak reached by women born around 1930) with the period and the age functions, is able to incorporate the cohort effect and eventually forecasts decreasing lung cancer (and implicitly, a decreasing future proportion of smokers among female cohorts).

In general, we expect that adjacent mortality rates tend to retain their rankings over time, whereas values at widely-separated values may cross over; for example, infant mortality for females in England and Wales in 1900 was twice the mortality rate at age 80, but by 2000, it was only one-third of the age 80 value, whereas the ranking of ages 79, 80 and 81 remained the same. The Bayesian model does permit such relationships between different ages to be included, in contrast to the more flexible Lee–Carter models and the more rigid APC ones. However, this does not imply that the Bayesian model necessarily produced better forecasts.

To assess the accuracy of the model forecasts as previously explained, we have forecast values for the period 1978-2007 based on trends between 1950 and 1977. The mean absolute error for malignant neoplasm of trachea, bronchus and lung in Table 1 shows the superior accuracy of the APC lung

Table 1.   Mean absolute error in log death rates by sex, forecast method, and cause of death when forecast rates compared with actual over period 1978-2007 (see text)

	Lung		IPB		MVA		Overall		
	Male	Female	Male	Female	Male	Female	Male	Female	*Total*
LC	0.88	0.45	0.80	0.63	0.85	0.95	0.22	0.17	*0.62*
BMS	0.65	0.44	0.85	0.69	0.39	0.77	0.22	0.15	*0.52*
Bayesian	0.30	0.51	0.93	0.81	0.40	0.89	0.24	0.15	*0.53*
APC	0.27	0.10	0.95	0.84	0.61	0.74	0.53	0.47	*0.56*
Average		*0.45*		*0.81*		*0.70*		*0.27*	
APC (35+)	–	–	0.72	0.65	0.61	0.94	0.19	0.11	

cancer forecast, especially for females — the difference among men between the APC and Bayesian models is small. The mean absolute error by age and time (table not shown) of the LC and BMS models decreases with age for males while it increases for females. This is directly related to the ability of models to incorporate non-linear trends. Since the models were fitted using data up to 1977, this means that for men, the observed trend is truncated just at the peak value among the oldest age groups (65 and over), while for women it occurs at the peak value among people aged 35-55. Clearly continuation of trends observed up to 1977 leads to an overestimation of future values among older men and middle-aged women.

As previously emphasised, the influenza, pneumonia, and bronchitis cause of death is mainly driven by period factors. If cohort patterns are partially pre-determined — for example, by earlier behaviours such as smoking — period effects, such as outbreaks of influenza, are likely to be less predictable in their timing. Results for the four models fitted over the period 1950-2007 for males are shown in Figure 4 (patterns for females are similar and are therefore omitted). The LC and BMS models fitted to out-of-sample data are more accurate than the APC model (Table 1). In practice, it appears that use of a drift value based on the average over the fitted time period seems to provide a reasonable model even for unpredictable changes in time. For the APC model a very poor fit is observed at young ages, mainly due to the extrapolation of the cohort function at 1970, which affects young cohorts.

Motor vehicle accident trends appear to be determined in a complex way (Figure 1, final panel). Factors such as public policies on traffic rules, alcohol or drug restrictions, investment in infrastructure and volume of road traffic all affect mortality. It is unclear if specific cohorts' behaviours affect the level of mortality (Jau-Yih *et al.*, 1996). The four models are characterised (left panel of Figure 5) by high levels of accuracy in forecasting the logarithm of death rates at ages 20-50 but with lower accuracy at other ages. A similar trend is observed in the female case for the LC and BMS models (right panel in Figure 5). The mean absolute error (MVA columns of Table 1) suggests that the BMS model has the highest level of forecast accuracy (although the BMS accuracy is matched by Bayesian model for males and the APC model for females).

The forecasts for the overall mortality log-rates in the period 1978-2007 (figure omitted) show no major differences between the first three models (LC, BMS, and Bayesian). The linearity in past trends makes the assumption of random walk with drift (on which LC and BMS are usually based) sufficient for an acceptably accurate forecast. The mean absolute error (Overall columns in Table 1) is similar for the LC and BMS models for males, although the BMS model performs slightly better than the LC model for females. The more complex Bayesian model does not show better results than the LC-family based models (Table 1). The APC forecast clearly has a higher mean absolute error than any other model; however the level of
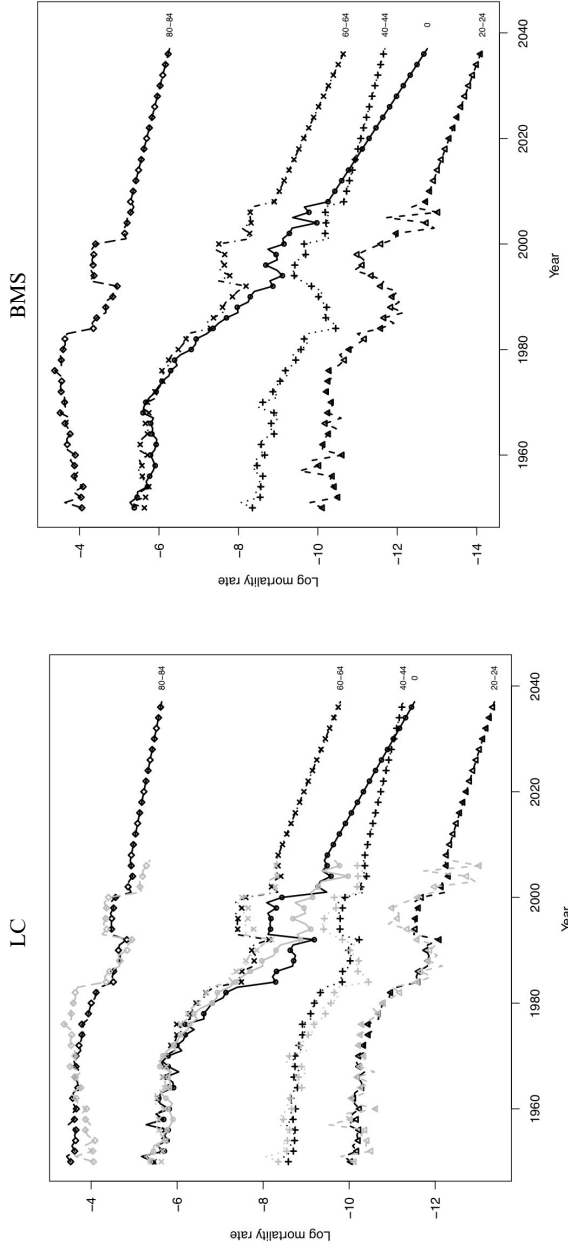
Notes: See Notes to Figure 2. Values at age zero not shown of APC model (see text).

Figure 4. Influenza, pneumonia, and bronchitis log-death rates by age, observed (light grey), fitted, and forecast values — LC, BMS, Bayesian, and APC model — Males
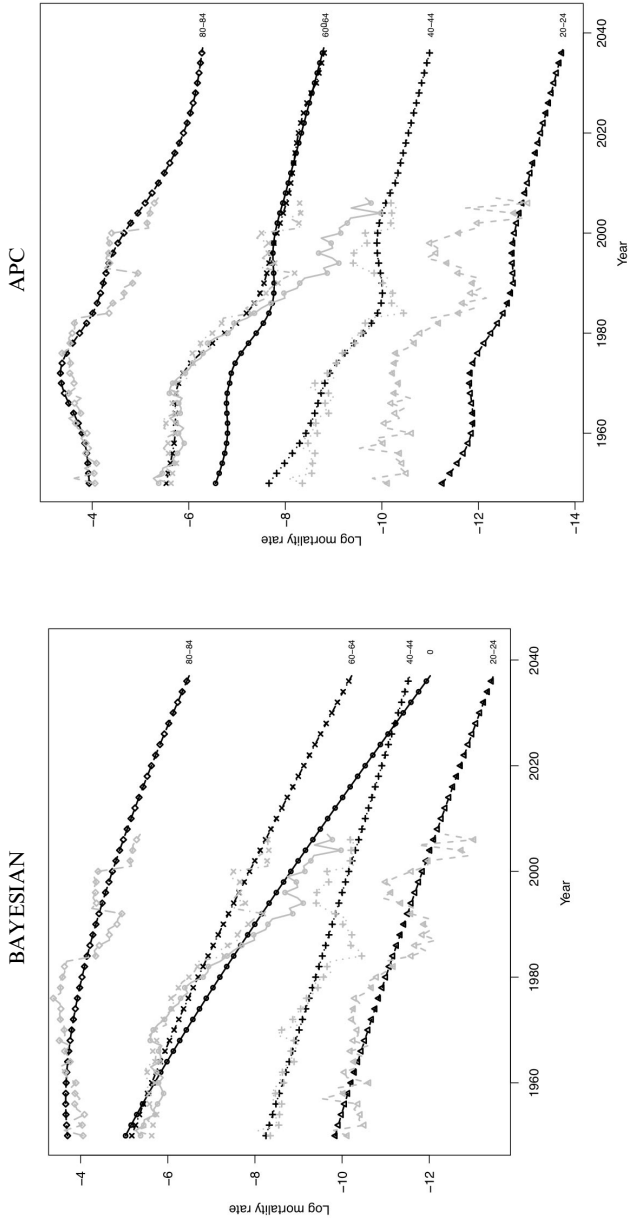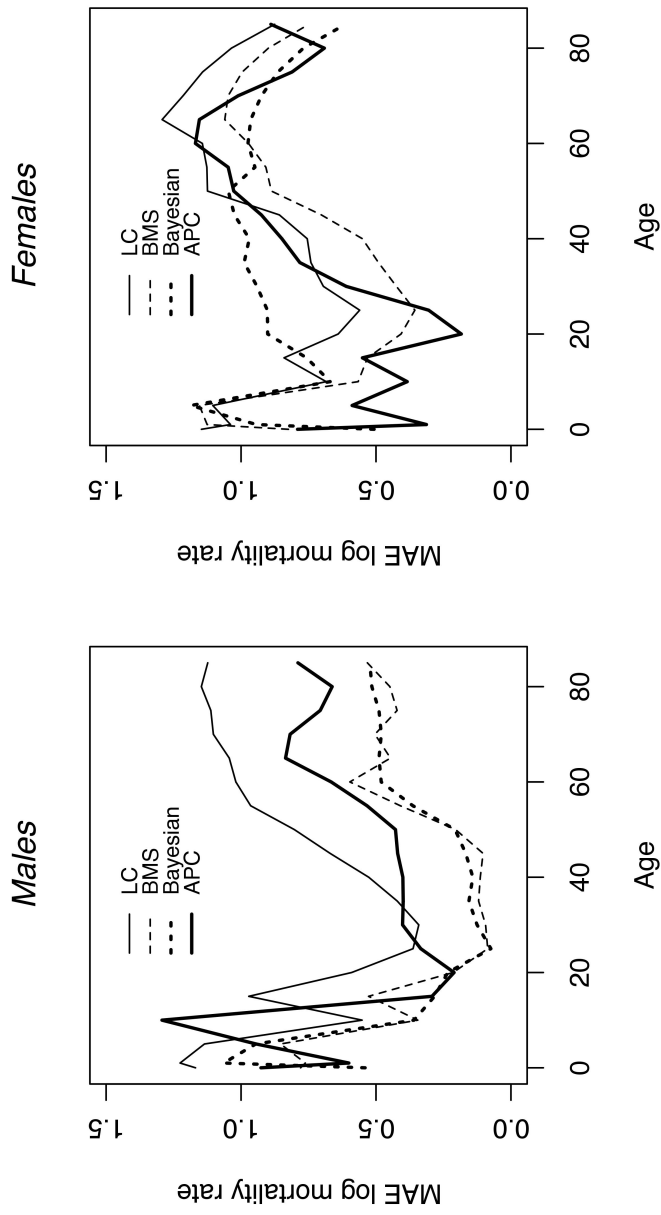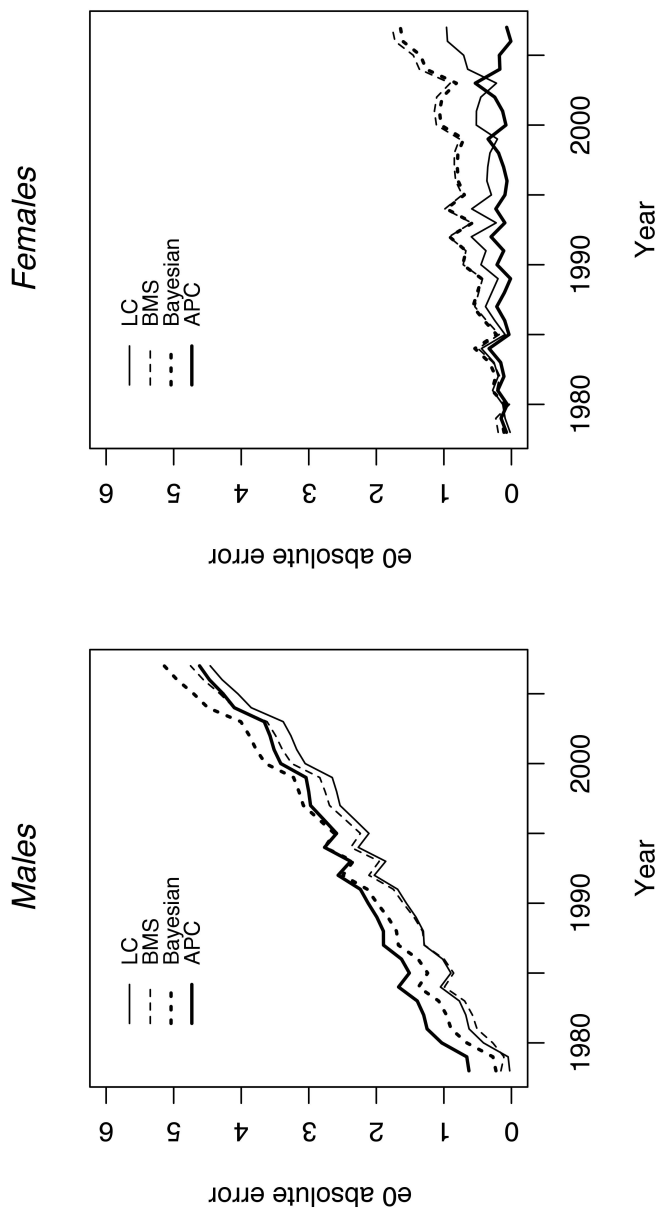
Figure 4. Continued

Figure 5.   Mean Absolute Error in log death rates by age, method, and sex — MVA

Note: all models are fitted to years 1950-1977 and MAE computed over years 1978-2007.

Note: all models are fitted to years 1950-1977.

Figure 6.   Absolute error in life expectancy at birth by sex and method — 1978-2007.

accuracy is similar to the other models if the mean absolute error is based only on age groups 35 and over.

The average accuracy across time over age and the average across age over time suggests improving accuracy among women aged 45 years and over, but such an improvement is not visible for males (results not shown). Among the three causes of death analysed, the forecasting models perform better in the case of lung cancer, with strong cohort-driven mechanism, and worst with the less predictable influenza-pneumonia and bronchitis case. However, the former case shows less variability and would be expected to be able to be forecast more accurately in any case.

Finally we present forecasts of life expectancy at birth in the period 1978-2007 using forecasts of overall mortality from the four models above fitted over the period 1950-1977. The absolute error over time (Figure 6) shows that for males, all four methods fail to forecast the improvement in their mortality experienced in the last three decades and therefore all underestimate projected life expectancy. For females, life expectancy at birth is far better forecast than for men, reflecting the fact that their mortality has followed a more stable trend over time. In the first years following the last observed points, the four methods are very similar (as all would be expected to follow the recent trend), but the longer term forecast shows somewhat better fits for the APC and the Lee–Carter models.

6. Discussion

The main goal of this paper has been to apply a range of alternative forecasting techniques to examples of causes of death with different underlying age and time patterns to assess which method copes better with the specificities of each case. Results show major differences among the three forecasting techniques — Lee–Carter and its Booth–Maindonald–Smith variant, Age-Period-Cohort model and Bayesian approach — and four main points can be drawn from this study.

Firstly, the Lee–Carter and Booth–Maindonald–Smith based on the random walk with drift, represents a valid option for forecasting cause of deaths characterised by linear trends. In situations in which the drivers of past trends act in a largely linear fashion (for example, economic circumstances and human capital might be expected to follow such a broad model on average), Lee–Carter based approaches can be considered the best choice due to their straightforward assumptions and limited need for subjective judgment. In addition, the Lee–Carter forecasting family copes reasonably well even with unpredictable changes in trends, including causes of deaths, such as motor vehicle accidents, and causes of death which are characterised by unpredictable period effects, such as for influenza-pneumonia-bronchitis.

Secondly, causes of deaths characterised by clear cohort patterns, such as lung cancer, are better-forecast using models that incorporate such an effect explicitly. The age-period-cohort model results show the best outcome for both male and female forecasts of trachea, bronchus and lung cancer.

Thirdly, the age-period cohort model, which estimates the period function together with the age and cohort functions, does not produce better forecasting results than the LC family methods when applied to period-driven cause of deaths.

Fourthly, while the Bayesian model produces good forecasts on average, it never shows significantly better results than the other models considered. This does not mean that the Bayesian approach is not a valid forecasting option but simply that in cases in which complete and reliable historical trends are available, no covariates are taken into consideration, and prior distribution is not relevant for the future forecasting, meaning the full benefits of this method are not realised.

In addition this study shows the need for analysing lung cancer (cohort-driven) mortality with models and techniques able to extract and forecast the cohort component separately from period and age factors. A major issue is to find the right way to forecast lung cancer due to the strong relation with smoking habits, smoking is responsible for about 90% of lung cancer deaths (Peto *et al.*, 1992), and to forecast the impact on health costs.

Table 1 shows that for these causes analysed, the rank ordering of the accuracy of forecasting performance for lung cancer is reversed compared with IPB, showing that a single method is inappropriate and the answer to the question of the best method depends on the nature and drivers of the time series (and models fitted over different periods could produce different results: past performance is no guarantee of performance in years to come). The analysis reported here is based on a limited number of causes of death for a single country for a specific time period. The extent to which the results generalise to other contexts awaits further work. There are arguments that overall mortality should be better forecast by combining disaggregated cause of death forecasts. However, the way that this is done is not specified. In this paper, we show that the examples of causes investigated here are better-modelled using a variety of approaches, so a disaggregated forecast would need to evaluate a range of alternatives in order to identify the most appropriate model in each case. This would be a major empirical exercise.
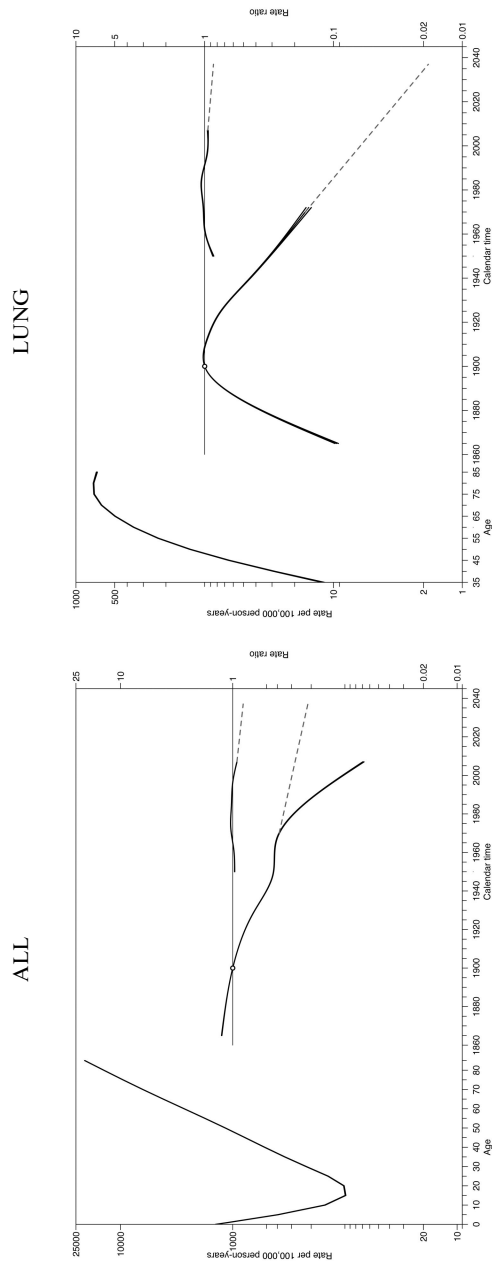
REFERENCES

BARBI, E. & VAUPEL, J.W. (2005). Comment on Inflammatory Exposure and Historical Changes in Human Life-Spans. *Science*, **308**(5729), 1743a.

BOOTH, H., MAINDONALD, J. & SMITH, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, **56**(3), 325-336.

BOOTH, H., HYNDMAN, R.J., TICKLE, L. & DE JONG, P. (2006). Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, **15**(9), 289-310.

BOOTH, H. & TICKLE, L. (2008). *Mortality modelling and forecasting: a review of methods.* ADSRI Working Paper, No. 3. Available at
http://adsri.anu.edu.au/pubs/ADSRIpapers/ADSRIwp-03.pdf

BROCK, A., GRIFFITHS, C. & ROONEY, C. (2006). The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales. *Health Statistics Quarterly*, **29**, 9-17.

CARSTENSEN, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, **26**, 3018-3045.

CATALANO, R. & BRUCKNER, T. (2006). Child mortality and cohort lifespan: a test of diminished entelechy. *International Journal of Epidemiology*, **35**(5), 1264-1269.

CLERIES, R., MARTINEZ, J.M., VALLS, J., PAREJA, L., ESTEBAN, L., GISPERT, R., MORENO, V., RIBES, J. & BORRÀS, J.M. (2009). Life expectancy and age-period-cohort effects: analysis and projections of mortality in Spain between 1977 and 2016. *Public Health*, **123**(2), 156-162.

CRIMMINS, E.M. (1981). The changing pattern of American mortality decline 1940-1977, and its implication for the future. *Population and Development Review*, **7**(2), 229-254.

DAVY, M. (2006). Time and generational trends in smoking among men and women in Great Britain, 1972-2004/5. *Health Statistics Quarterly*, **32**, 35-43.

DE JONG, P. & TICKLE, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, **13**(1), 1-18.

FINCH, C.E. & CRIMMINS, E.M. (2004). Inflammatory exposure and historical changes in human life-spans. *Science*, **305**, 1736-1739.

GIROSI, F. & KING, G. (2008). *Demographic forecasting*. Princeton University Press.

GRIFFITHS, C. & BROCK, A. (2003). Twentieth Century mortality trends in England and Wales. *Health Statistics Quarterly*, **18**, 5-17.

GRIFFITHS, C. & ROONEY, C. (2006). Trends in mortality from Alzheimer's disease, Parkinson's disease and dementia, England and Wales, 1979-2004. *Health Statistics Quarterly*, **30**, Summer, 6-14.

HEATHCOTE, C. & HIGGINS, T. (2001). A regression model of mortality, with application to the Netherlands. In: E. TABEAU, A. VAN DEN BERG JETHS & C. HEATHCOTE (eds.). *Forecasting mortality in developed countries*. Kluwer Academic Publishers.

HOLLMANN, F.W., MULDER, T.J. & KALLAN, J.E. (2000). Methodology and assumptions for the population projections of the United States: 1999 to 2100. *Population Division Working Paper No. 38*, U.S. Bureau of the Census.

HYNDMAN, R.J. & ULLAH, M.S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics and Data Analysis*, **51**(10), 4942-4956.

JAU-YIH, T., WEN-CHUNG, L. & JUNG-DER, W. (1996). Age-period-cohort analysis of motor vehicle mortality in Taiwan, 1974-1992. *Accident Analysis and Prevention*, **28**(5), 619-626.

LEE, R.D. & CARTER, L.R. (1992). Modelling and forecast U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659-671.

LEE, R.D. & MILLER, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, **38**(4), 537-549.

Lutz, W. & Goldstein, J.R. (2004). *How to deal with uncertainty in population forecasting?* IIASA Reprint Research Report, RR-04-09. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mason, W.M. & Smith, H.L. (1985). *Age-Period-Cohort analysis and the study of deaths from pulmonary tuberculosis*. In: W.M. Mason & S.E. Fienberg (eds.). *Cohort analysis in social research*. Springer-Verlag, New York.

Murphy, M. (2010). Re-examining the dominance of birth cohort effects on mortality. *Population and Development Review*, **36**(2), 365-390.

Office for National Statistics (2000). *20th century mortality* (England & Wales 1901-2000) CD-ROM.

Office for National Statistics (2009). *21st century mortality database*. Available at http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=6922

Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics*, **7**(4), 530-550.

Peto, R., Lopez, A.D., Boreham, J., Thun, M. & Heath, C. (1992). Mortality from tobacco in developed countries: indirect estimation from national vital statistics. *Lancet*, **339**, 1268-1278.

Polder, J.J., Barendregt, J.J. & van Oers, H. (2006). Health care costs in the last year of life — the Dutch experience. *Social Science and Medicine*, **63**(7), 1720-1731.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Available at http://www.R-project.org

Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**, 556-570.

Scitovsky, A.A. (1994). The high cost of dying revisited. *Milbank Quarterly*, **72**(4), 561-591.

Tabeau, E. (2001). *A review of demographic forecasting models for mortality*. In: E. Tabeau, A. van den Berg Jeths & C. Heathcote (eds.). *Forecasting mortality in developed countries*. Kluwer Academic Publishers.

Tabeau, E., Ekamper, P., Huisman, C. & Bosch, A. (1999). Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends. *European Journal of Population*, **15**, 153-183.

Yang, Y., Schulhofer-Wohl, S., Fu, W.J. & Land, K.C. (2008). The intrinsic estimator for Age-Period-Cohort analysis: what it is and how to use it. *American Journal of Sociology*, **113**(6), 1697-1736.

Forecasting Mortality, Different Approaches for Cause of Deaths?

APPENDIX

Figure A1.   Age-Period-Cohort factors observed (solid line) and forecast (dashed line) values by cause of death — Males
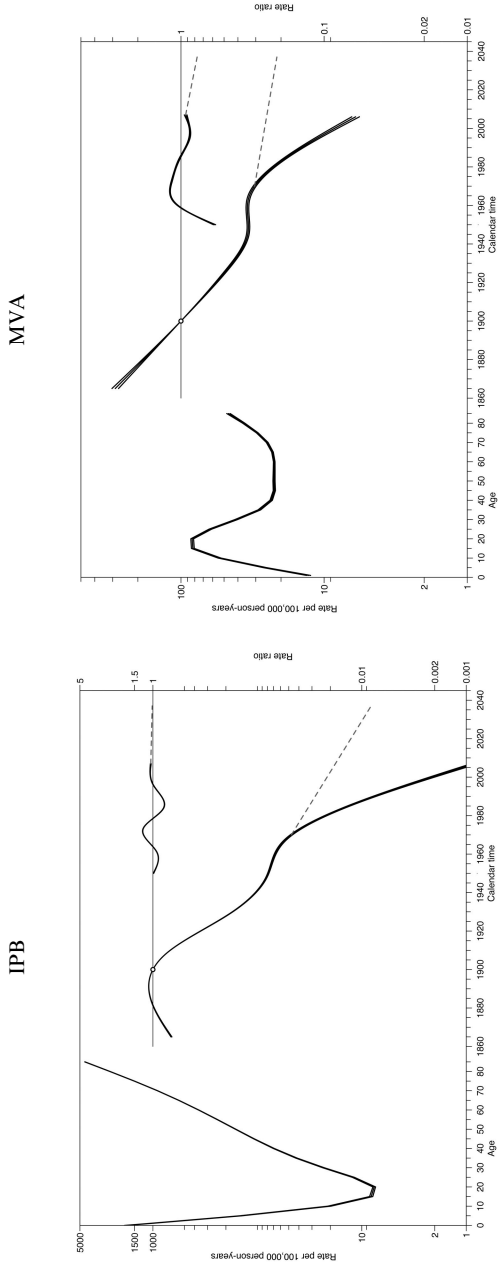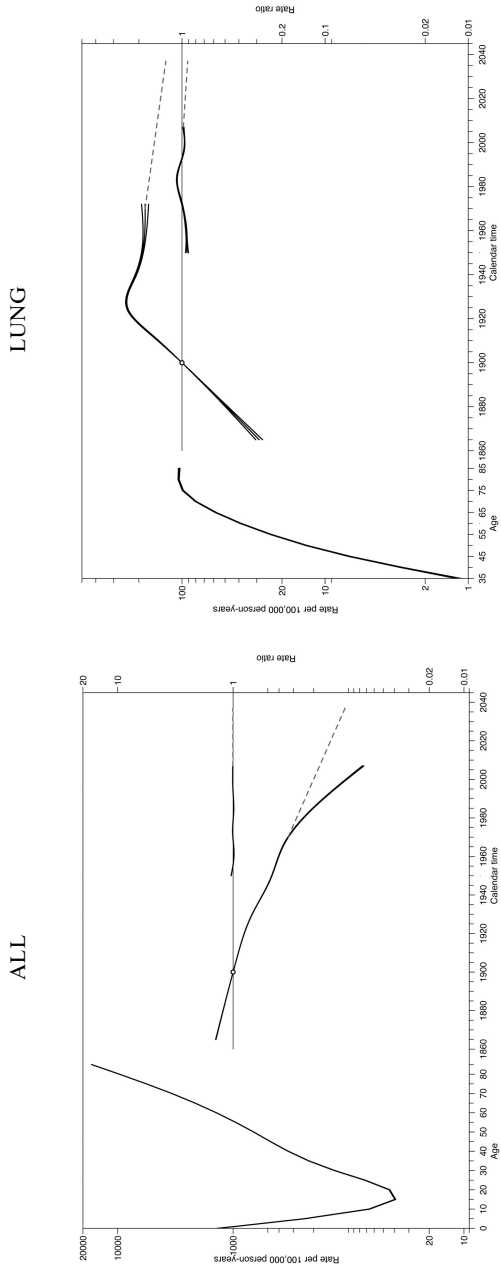
Figure A1. Continued

Note: all models are fitted to years 1950-2007.

Figure A2.   Age-Period-Cohort factors observed (solid line) and forecast (dashed line) values by cause of death — Females
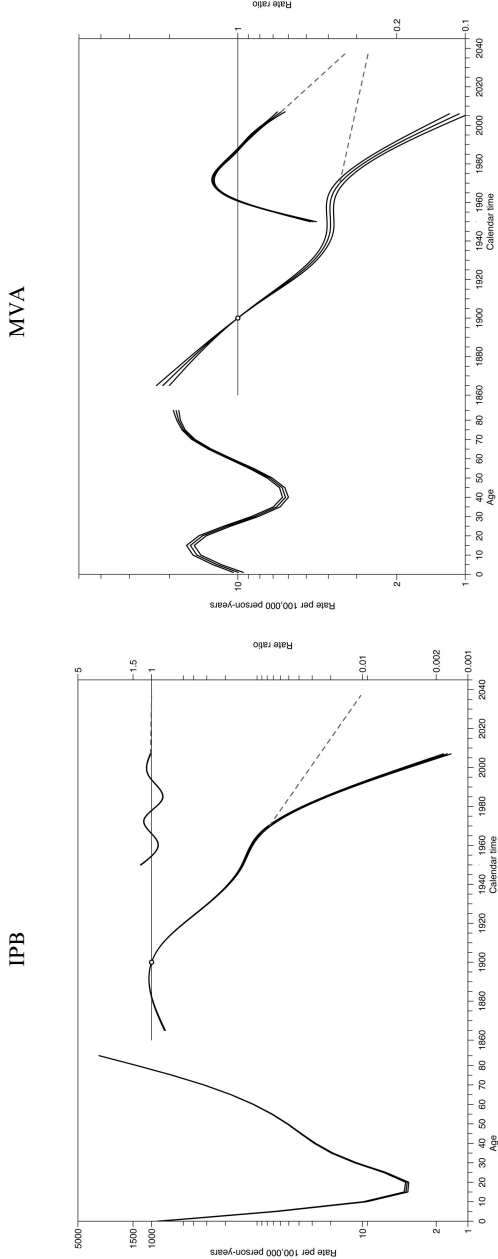
Figure A2.   Continued