Lung cancer rate predictions using generalized additive models

MARK S. CLEMENTS*
National Centre for Epidemiology and Population Health, The Australian National University, Canberra, ACT 0200, Australia
Mark.Clements@anu.edu.au

BRUCE K. ARMSTRONG
School of Public Health, The University of Sydney, Sydney, Australia

SURESH H. MOOLGAVKAR
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

SUMMARY
Predictions of lung cancer incidence and mortality are necessary for planning public health programs and clinical services. It is proposed that generalized additive models (GAMs) are practical for cancer rate prediction. Smooth equivalents for classical age-period, age-cohort, and age-period-cohort models are available using one-dimensional smoothing splines. We also propose using two-dimensional smoothing splines for age and period. Variance estimation can be based on the bootstrap. To assess predictive performance, we compared the models with a Bayesian age-period-cohort model. Model comparison used cross-validation and measures of predictive performance for recent predictions. The models were applied to data from the World Health Organization Mortality Database for females in five countries. Model choice between the age-period-cohort models and the two-dimensional models was equivocal with respect to cross-validation, while the two-dimensional GAMs had very good predictive performance. The Bayesian model performed poorly due to imprecise predictions and the assumption of linearity outside of observed data. In summary, the two-dimensional GAM performed well. The GAMs make the important prediction that female lung cancer rates in these countries will be stable or begin to decline in the future.

Keywords: Age-period-cohort model; Bayesian; Lung cancer; Trends.

1. INTRODUCTION
Predictions of future cancer incidence and mortality rates, defined here as rate projections, are necessary for planning public health programs and clinical services, including clinical training (Hakulinen, 1996). As a motivating example, lung cancer is the second most common site for cancer incidence and the most common site for cancer deaths in the United States (Ries et al., 2003).

*To whom correspondence should be addressed.
One challenge for the prediction of lung cancer rates is the complex and rapid pattern of change over
the past century. For the United States, lung cancer incidence rates for males reached a peak during the
mid 1980s, while rates for females may have reached a peak by the late 1990s (Weir et al., 2003). The
rates display a strong association with age and the age pattern has changed between cohorts and between
periods. Such rate patterns can largely be explained by marked changes in smoking behavior (Burns et al.,
1997). Importantly, lung cancer rates are associated with cumulative smoking exposure, so that future
lung cancer rates are predictable and are expected to follow a smooth curvilinear pattern. However, there
is a paucity of historical smoking data, so that lung cancer predictions are commonly based solely on
mortality or incidence rates.

Simple rate models, such as the age-drift (Clayton and Schifflers, 1987a) and linear power models
(Dyba et al., 1997; Møller et al., 2003) provide useful mechanisms for rate projections of linear or log-
linear trends, particularly for rare cancers. Modeling issues arise when data exhibit curvilinear trends or
age-period-cohort interactions. For modeling curvilinear trends, alternative approaches include the use of
polynomial terms, which may suffer from rapid changes at the ends, or regression splines, which require
the specification of knots (Heuer, 1997). For age-period-cohort interactions, classical age-period-cohort
models are well suited to investigating the descriptive epidemiology for a disease (Holford, 1983; Clayton
and Schifflers, 1987b). However, they are less suitable for cancer projections, requiring specification of the
parametric forms for projected period and cohort effects outside of observed data (Osmond, 1985; Møller
et al., 2003). Additionally, classical models often use data in 5-year periods and 5-year age groups to
approximate the following cohorts through time, hence aggregating across data that are typically available
by single-year periods and 5-year age groups.

Smooth changes in rates suggest the use of a continuous time rate modeling approach using modern
regression methods (Keiding, 1990). Such models can be interpreted geometrically as fitting a smooth
surface through rates over the Lexis diagram, where projections are estimated by extending the rate surface
out in time. Models to implement this approach include kernel smoothing (Keiding, 1990), local likeli-
hood regression (Loader, 1999), and generalized additive models (GAMs) (Hastie and Tibshirani, 1990;

GAMs have been used to describe trends in age-standardized cancer rates (Boyle et al., 2003) and for
back-calculation of HIV incidence using two-dimensional smoothers (Marschner and Bosch, 1998). These
models have received little attention for modeling age-period-cohort cancer data, although the application
has been suggested by several authors (Bashir and Estève, 2001; Berzuini and Clayton, 1994; Heuer,
1997). The limited attention may relate to two technical hurdles. First, the models require the specifica-
tion of smoothing parameters, where previous implementations have either required the user to investigate
manually the possible values using cross-validation (Hastie and Tibshirani, 1990) or have been computa-
tionally expensive (Gu, 2002). Second, an analysis of age-period-cohort data typically requires variance
estimates for weighted sums of correlated estimates, such as age-standardized rates and total population
counts (Hakulinen and Dyba, 1994), demanding that variance estimation use either linear approximations
such as the delta method or resampling methods such as the bootstrap.

Fortunately, efficient methods for automatic smoothing parameter estimation for thin-plate smoothing
splines have recently been developed and implemented (Wood, 2000, 2003). Moreover, intensive calcu-
lations such as bootstrapping are now feasible. Given these recent developments, we propose that GAMs
provide a useful mechanism for estimating rate projections.

In the remainder of the paper, data sources are first described, then methods are introduced for
using GAMs for rate projections, including variance estimation using the bootstrap. The models
will be compared with a Bayesian age-period-cohort model, with model comparisons using both cross-
validation and measures of predictive performance. Following an application of the methods to lung
cancer mortality data from the World Health Organization (WHO) Mortality Database, there is a brief
discussion.
2. DATA SOURCES

The primary data source was the WHO Mortality Database with URL http://www3.who.int/whosis/menu.cfm?path=mort. Data were extracted for population estimates and for cancer deaths for the lung, bronchus, and trachea. Data were available by country, by single calendar year, and by 5-year age groups for females aged 25–84 years. The countries included in the analysis were Australia (1952–2001), New Zealand (1951–2000), Sweden (1952–2001), United Kingdom (1950–1999), and the United States (1951–2000).

Age and period were represented by the midpoint of an age group and a period interval, respectively. Birth cohorts were calculated by the difference between period and age. For age standardization, the revised WHO world population was used as the standard population (Ahmad et al., 2002).

3. GENERALIZED ADDITIVE MODELS

Let the observed number of event counts be represented by $y$, with age represented by $a$, period by $p$, and cohort by $c = p - a$. Following Brillinger (1986), we assumed that the observed event counts followed a Poisson distribution. Let the person-years of exposure estimated by the annual population be represented by $n$. Moreover, let the predicted mean for the number of event counts be $\mu$. Let $\{w_a\}$ be a set of positive-valued standard weights indexed over age groups $a$, such that $\sum_a w_a = 1$. For a given period $p$, the predicted age-standardized rate will be

$$\text{ASR}_p = \sum_a w_a \frac{\mu_{ap}}{n_{ap}}.$$

In order to model the rate, the logs of the person-years of exposure were included as offsets in the regression model. In the following, let $g_j(x)$ and $g_j(x, y)$ represent one-dimensional and two-dimensional smoothing functions, respectively.

The proposed approach is to use a class of age-period-cohort models using GAMs. Let Model A + P be a GAM with smoothed age and period, with the expected value following the relationship $\log(\mu_{ap}) = \log(n_{ap}) + g_1(a) + g_2(p)$. Similarly, let Model A + C be a GAM smoothing for age and cohort, defined by $\log(\mu_{ac}) = \log(n_{ac}) + g_1(a) + g_2(c)$, and let Model A + P + C be a GAM smoothing for age, period, and cohort, defined by $\log(\mu_{apc}) = \log(n_{apc}) + g_1(a) + g_2(p) + g_3(c)$.

The definition for cohort is based on an age interval of 5 years and a period interval of 1 year. Holford (1983) described a ‘saw-tooth’ or cyclic pattern that can be generated from age-period-cohort modeling of data based on unequal intervals. To resolve this problem, we can assume smoothness within the 5-year age interval and fit splines or another smoother for cohort and period effects (Heuer, 1997).

Models A + P and A + C are smooth equivalents of the age-period and age-cohort models described by Clayton and Schifflers (1987a), which assume proportionality of the age-specific rates across periods and cohorts, respectively. Model A + P + C is a smooth equivalent of the age-period-cohort model described by Clayton and Schifflers (1987b). Finally, let Model A * P be a GAM defined using a two-dimensional smoothing function, such that $\log(\mu_{ap}) = \log(n_{ap}) + g_1(a, p)$. For flexible determination of the smoothing, the models will be estimated using penalized likelihood. Fitting of the models involves determination of one or two smoothing parameters together with the spline parameters. Thin-plate smoothing splines were used for the functions $g_j$ (Green and Silverman, 1994; Wood, 2003). We want to find the function $g$, dependent upon the unpenalized likelihood $l(-)$, unknown parameters $\theta$, and smoothing parameters $\lambda_j$, that maximizes the penalized log-likelihood

$$l(\theta) - \frac{1}{2} \sum_j \hat{\lambda}_j J(g_j),$$
with the summation indexed by $j$ over the smoothers. For univariate thin-plate splines, $J(g) = \int_\mathbb{R} g''(t)^2 dt$. The resulting thin-plate spline takes the form

$$g(t) = \sum_i \delta_i |t - t_i|^3 + b_1 + b_2 t,$$

where $\delta_i$ and $b_k$ are constants, subject to the identifiability constraints that $\sum_i \delta_i = \sum_i \delta_i t_i = 0$.

Note that these splines are equivalent to natural cubic smoothing splines. Outside of observed data, $g''(t) = 0$, hence, $g(t)$ and $\log(\mu(t))$ will be linear. As a consequence, predictions for a given age from Models A + P, A + C, and A + P + C will be linear on a log scale for later periods and cohorts.

For the two-dimensional thin-plate smoothing spline,

$$J(g) = \int \int_\mathbb{R}^2 \left[ \left( \frac{\partial^2 g}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 g}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 g}{\partial y^2} \right)^2 \right] dx dy,$$

so that, writing $t = [x, y]^T$, the thin-plate spline then takes the form

$$g(t) = \sum_i \delta_i \eta(\|t - t_i\|) + (b_1 + b_2 x + b_3 y),$$

where $\delta_i$ and $b_k$ are constants, $\|\cdot\|$ denotes the Euclidean norm,

$$\eta(r) = \begin{cases} \frac{1}{16\pi} r^2 \log(r^2), & \text{for } r > 0, \\ 0, & \text{for } r = 0, \end{cases}$$

and we impose the identifiability constraints that $\sum_i \delta_i = \sum_i \delta_i x_i = \sum_i \delta_i y_i = 0$.

For fixed $x$, such as for a given age, we note that $g(t)$ as a function of $y$ will have curvature that is between that for $y$ and $y^2 \log(y)$. Moreover, $g$ will tend to be linear in $y$ when $y$ is large because $\lim_{y \to \infty} \partial^2 g/\partial y^2 = 0$. Predictions from these models are globally based, in the sense that all of the observed data are used in estimating the predictions.

The models were implemented using the mgcv package (Wood, 2000, 2003) in R (Ihaka and Gentleman, 1996). The smoothing parameters $\lambda_j$ were automatically estimated using unbiased risk estimation, which is equivalent to minimizing Akaike’s information criterion. Similar results were obtained when the smoothing parameters were estimated using generalized cross-validation. Given potential concerns about numerical stability, we rescaled age, period, and cohort by a factor of 0.001 and found consistent parameter estimates. Attention to the default convergence criterion may be required when applying the gam package by Trevor Hastie.

### 3.1 Bootstrap estimation

As estimated total counts and age-standardized rates are a weighted sum of correlated age-specific rates, interval estimation should take the correlation structure into account (Hakulinen and Dyba, 1994). The bootstrap provides a flexible method for estimation of confidence intervals for the mean and prediction intervals for the GAMs (Davison and Hinkley, 1997). The parametric bootstrap has been used previously for estimating confidence intervals for age-period-cohort models (Robertson and Boyle, 1998).

In outline, a nonparametric bootstrap was implemented using resampling of model-based residuals with an adjustment for bias due to smoothing. The smoothing parameters were reestimated at each bootstrap iteration, taking some account of uncertainty in the parameters (Hastie et al., 2001, p. 231). The data were initially fitted with smoothing parameters estimated by minimizing the unbiased risk estimator.
The simulated outcomes were calculated by resampling residuals from an undersmoothed fit added to oversmoothed means. The under- and oversmoothing take some account of the bias-precision trade-off inherent in nonparametric regression.

Let the predicted means from the initial fit be \( \hat{\mu}_i \) with smoothing parameters \( \hat{\lambda}_j \). Let \( \tilde{\mu}_i \) be the mean from an oversmoothed fit using smoothing parameters \( 2\hat{\lambda}_j \) and let \( \hat{\mu}_i \) be the mean from an undersmoothed fit using smoothing parameters of \( \hat{\lambda}_j/2 \). The factor of 2 in the over- and undersmoothing is as suggested by Davison and Hinkley (1997, p. 365).

For each simulation, the Pearson standardized residuals \( r_i \) can be defined using the undersmoothed means and hat values \( h_i \), such that \( r_i = (y_i - \tilde{\mu}_i)/\sqrt{\hat{\mu}_i(1 - h_i)} \) are scaled to have mean zero. The simulated responses \( y_i^* \) combine the oversmoothed mean and the resampled residuals \( r_i^* \) such that \( y_i^* = \max[0, \tilde{\mu}_i + r_i^*\sqrt{\hat{\mu}_i}] \).

The model was fitted to the simulated responses using an unbiased risk estimation, including estimation of the smoothing parameter at each simulation. Mean predictions for covariates \( x \) were obtained from \( \exp[\log \hat{\mu}(x) + \log \tilde{\mu}_i^*(x) - \log \hat{\mu}_i(x)] \).

Heteroscedasticity of residuals required stratified resampling of residuals for expected counts less than 10 (Davison and Hinkley, 1997). For the \( A + P \) and \( A + C \) models, which often had poor model fits and large estimated residuals, the bootstrap simulations were based on the \( A \ast P \) model. Similar results were obtained when we used standardized deviance residuals. The bootstrap was implemented in R using the boot package written by Angelo Canty and described by Davison and Hinkley (1997). Three hundred bootstrap iterations were used.

### 4. A Bayesian age-period-cohort model

We would like to compare the predictive performance of GAMs with the performance of a model standard. Although Møller et al. (2003) provide a thorough review of model classes for rate prediction, we would expect that most of the additive models presented would perform poorly with lung cancer rate data because they take no account of age-period-cohort interactions. Alternatively, the Bayesian age-period-cohort rate model proposed by Berzuini and Clayton (1994) is of interest for three reasons. First, the model provides an alternative smooth analog to the classical age-period-cohort models, taking account of potential interactions between the three time parameters. Second, Bray (2002) found that the model performed well compared with additive models for a range of cancers, based on plug-in estimates of deviance for validation data. Third, the Bayesian model formulation suggests a predictive approach to model selection, where parameter uncertainty can be included in model comparisons (Gelman et al., 1995). In practice, this involves averaging a model fit parameter over the posterior distribution. Such an approach would be useful in a likelihood-based setting, possibly through the use of the bootstrap (see Hastie et al., 2001, p. 235).

The Bayesian age-period-cohort model is a generalization of parametric age-period-cohort modeling, using an autoregressive smoothing of the separate effects of age, period, and cohort. The formulation proposed by Berzuini and Clayton (1994) used second-order autoregressive smoothing, with effects for cohort and period being linear outside of observed data. For a discussion of the autoregressive priors, see Bashir and Estève (2001) and Bray (2002).

An adjustment of the model proposed by Berzuini and Clayton (1994) was required for implementing the cross-validation. For each iteration of the cross-validation, there may be no data within a stratum for an age, a period, or a cohort, suggesting the use of an undirected prior for early effects. If we let \( \{\beta_b, b = 1, \ldots, B\} \) represent the parameters for age, period, or cohort, then we assumed that

\[
\beta_1|\beta_2, \beta_3 \sim N(2\beta_2 - \beta_3, \sigma_2^2), \quad \beta_2|\beta_1, \beta_3 \sim N\left(\frac{\beta_1 + \beta_3}{2}, \sigma_2^2\right), \\
\beta_b|\beta_1, \ldots, \beta_{b-1} \sim N(2\beta_{b-1} - \beta_{b-2}, \sigma_2^2), \quad b > 2.
\]
Rate predictions using generalized additive models

Predictions were calculated by pushing out the period and cohort terms using the same priors. Gamma-distributed priors were assigned to the precision parameters, with shape and scale parameters of 0.001. To assess the choice of these priors, we also used gamma-distributed priors with shape and scale parameters of 0.5 and 0.0005, respectively, consistent with a large mean and variance, and found very similar values for the means and fit statistics. The model was implemented using WinBUGS interfaced with R. For the main model fit, we used a Monte Carlo Markov chain (MCMC) length of 10000. The interface allowed the use of existing functions for cross-validation, averaging two sets of 10-fold validation with a chain length for each MCMC fit of 2000 (see next section).

5. Model comparisons and predictive performance

Our interest is in comparing GAMs fitted using maximum likelihood with a Bayesian age-period-cohort model where the posterior distribution is evaluated using MCMC methods. Predictive performance based on out-of-data model validation for cancer predictions was discussed recently by Møller et al. (2003). The authors defined recent short-term predictions as fitting models to observed data excluding the last 10 years of observation, with the predictions being for the last 5 years of observation.

We define the deviance as

\[ D(y, \mu) = -2 \log[p(y|\theta)] + 2 \log[p(y|\mu = y)], \]

where \( p(\cdot) \) denotes probability, \( \theta \) is the set of model parameters, and \( \mu = E(y) \), hence \( \mu = \mu(\theta) \). For independent Poisson observations indexed by \( i \), the deviance is

\[ D(y, \mu) = -2 \sum_i [(y_i - \mu_i) - y_i \log(y_i/\mu_i)], \]

with the convention that \( 0 \log 0 = 0 \). If all of the data are used for prediction, with the estimator denoted \( \hat{\mu} = \mu(E(\theta|y)) = \mu(\hat{\theta}) \), then the plug-in deviance for fitted data can be defined as \( D(y, \hat{\mu}) \). This definition for deviance is similar to that used for classical GAMs. If the data are divided into a training set, denoted by subscript \( t \), and an assessment (or test) set, denoted by subscript \( a \), then the plug-in deviance for assessment data can be defined as \( D(y_a, \hat{\mu}_t) \), where \( \hat{\mu}_t = \mu(E(\theta|y_t)) \). Note that the plug-in approaches make no adjustment for parameter uncertainty.

As is natural in a Bayesian setting, the predictive approach takes expectations over the posterior distribution, such that the expected deviance for fitted data is defined as \( E_{\theta|y} D(y, \mu(\theta)) \). Similarly, the expected deviance for assessment data, defined as \( E_{\theta|y} D(y_a, \mu(\theta)) \), provides an out-of-data measure of predictive performance, taking account of parameter uncertainty.

Information theoretic criteria related to Akaike’s information criterion take the form \( D(y, \hat{\mu}) + 2p^* \), where \( p^* \) is the number of model parameters. Such criteria are asymptotically equivalent to cross-validation and historically had not been used for Bayesian models. Spiegelhalter et al. (2002) recently described such an approach for Bayesian models, defining a measure \( p_D \) that provides an estimate of the number of parameters being the difference between the expected and plug-in deviance for fitted data, such that

\[ p_D = E_{\theta|y} [D(y, \mu(\theta))] - D(y, \mu(E(\theta|y))) = D(y, \mu) - D(y, \mu(\hat{\theta})). \]

The authors also introduced the Deviance Information Criterion (DIC), defined as

\[ DIC = D(y, \hat{\mu}) + 2p_D. \]

K-fold cross-validation is a related approach, where the data are randomly partitioned into \( K \) sets of similar size, indexed by \( k \), with model fitting on all of the data outside of set \( k \) together with an assessment
on set \( k \). Typically, \( K \) takes values such as 5 or 10. If \( \hat{\mu}_{-k} \) is defined as the estimator based on fitting the model on data outside of set \( k \), \( \hat{\mu}_{-k,k} \) is \( \hat{\mu}_{-k} \) restricted to set \( k \), and \( y_k \) are the outcomes in set \( k \), and \( p_k \) is the proportion of the data in set \( k \), then the \( K \)-fold cross-validation deviance with bias correction is defined as

\[
CV_K = \sum_{k=1}^{K} D(y_k, \hat{\mu}_{-k,k}) + D(y, \hat{\mu}) - \sum_{k=1}^{K} p_k D(y, \hat{\mu}_{-k})
\]

(Davison and Hinkley, 1997, p. 295). Note that \( K \)-fold cross-validation can be computationally expensive, requiring that the model be refitted \( K \) times. The theoretical properties for cross-validation are better understood than the DIC, so that cross-validation was considered the primary measure of fit for within-data predictions.

In the following, the models will be assessed using cross-validation, DIC and expected deviance for assessment data. The intention is to compare the use of cross-validation with the DIC, and to compare cross-validation with the expected deviance for assessment data.

### 5.1 Bootstrap estimation

The bootstrap can be used to apply a predictive approach to frequentist models. A bootstrap estimate for \( p_D \) can be calculated using

\[
p_D = E^*[D(y^*, \mu(E^*(\theta|y^*))) - D(y^*, \mu(\theta|y^*))],
\]

where \( E^*(\cdot) \) denotes expectation under the bootstrap. This approach is motivated by the bootstrap estimation of the extended information criterion (EIC) described in Konishi and Kitagawa (1996). Similarly, for assessment data, we are interested in a bootstrap estimator with lower variance than the naïve bootstrap estimator \( E^*[D(y^*_a, \mu(\theta|y^*_a))] \). By noting that

\[
E^*[D(y^*_a, \mu(\theta|y^*_a))] - D(y^*_a, \hat{\mu}_a) \approx E^*[D(y^*_a, \mu(E^*(\theta|y^*_a))) - D(y^*_a, \mu(\theta|y^*_a))],
\]

then a bootstrap estimate for the expected deviance for assessment data can be calculated using

\[
D(y^*_a, \hat{\mu}_a) + E^*[D(y^*_a, \mu(E^*(\theta|y^*_a))) - D(y^*_a, \mu(\theta|y^*_a))].
\]

The bootstrap simulations for \( y^*_a \) and \( y^*_t \) were based on fitting a model to all of the observed data. As an empirical validation, we found that estimates of \( p_D \) were similar to the estimated degrees of freedom from the trace of the estimated smoother matrix, and that the means for the expected deviance for assessment data were similar to that for the naïve bootstrap estimator.

### 5.2 Model assessment using data with different aggregations

Cancer mortality data are commonly reported by single-year calendar periods and 5-year age groups, however, the Bayesian models require age and period to be available by similar lengths of time, such as 5-year groups, in order for discrete changes in cohort to be sensibly defined. For assessment between models, models were fitted to the data using model-specific data aggregation, and then predictions and deviance were estimated for data by 5-year periods and 5-year age groups.

For the GAMs, predictions were based on bootstrapping and model estimation for single-year calendar periods, with deviance estimates based on results aggregated up to 5-year calendar periods. In particular, the estimated degrees of freedom \( p_D \) was calculated from the aggregated results. For comparison of cross-validation deviance between models, the GAMs were sampled using blocks of 5-year calendar periods.
Rate predictions using generalized additive models

by 5-year age groups, with the models fitted for single-year calendar periods, and estimates for model assessment used aggregated results.

6. Results

6.1 Descriptive epidemiology

Lung cancer mortality rates for U.S. females across the Lexis diagram are presented in Figure 1. The peak rates at given ages, represented by vertical lines, varied by period and by cohort. For future trends, most of the younger age groups are expected to exhibit declining rates, while the rates at oldest ages could conceivably rise further before reaching their ultimate peak.

6.2 Model fits

The fits for the different models are summarized in Table 1. As expected from classical analyses of lung cancer rates using age-period-cohort models, the age-cohort models performed better than the age-period models (Clayton and Schifflers, 1987a). The age-period and age-cohort models generally had worse model fits than the other three models, with Sweden as a notable exception.

Model selection based on cross-validation and the DIC did not strongly support either of the age-period-cohort models or the A*P model. For recent short-term projections, the plug-in deviances for GAM A + P + C and A*P were less than those for the Bayes A + P + C for most comparisons, except for Sweden. This indicates that the mean predictions were closer to the observed data for the GAMs. Taking account of parameter uncertainty using the predictive deviance for recent short-term projections, the GAMs performed considerably better for all comparisons, including models for Sweden. For the GAMs, the predictive deviance was lower for the A* P model for four of the five comparisons, where the GAM A + P + C model was better for Sweden.

For males (results not shown), it was found that the A*P model had the lowest predictive deviance for recent short-term predictions across all five countries.

6.3 Rate projections

For a graphical assessment of the models, age-standardized rates based on models fitted to data excluding the last 10 years of data were compared with observed data (see Figure 2). Model predictions and 95%
<table>
<thead>
<tr>
<th>Country</th>
<th>Model</th>
<th>Plug-in deviance</th>
<th>$p_D$</th>
<th>DIC</th>
<th>CV$_{10}$</th>
<th>Plug-in deviance</th>
<th>Pred deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P</td>
<td>46 686</td>
<td>26</td>
<td>46 737</td>
<td>71 614</td>
<td>60 809</td>
<td>60 769</td>
</tr>
<tr>
<td></td>
<td>GAM A + C</td>
<td>58 577</td>
<td>34</td>
<td>59 255</td>
<td>13 478</td>
<td>17 175</td>
<td>17 084</td>
</tr>
<tr>
<td></td>
<td>GAM A + P + C</td>
<td>433</td>
<td>42</td>
<td>51 847</td>
<td>10 542</td>
<td>10 333</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A * P</td>
<td>114</td>
<td>88</td>
<td>28 979</td>
<td>8 222</td>
<td>755</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bayes A + P + C</td>
<td>404</td>
<td>38</td>
<td>48 029</td>
<td>7 207</td>
<td>17 010</td>
<td></td>
</tr>
<tr>
<td>U.S.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recent projections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P</td>
<td>13 063</td>
<td>20</td>
<td>13 103</td>
<td>18 869</td>
<td>8 920</td>
<td>8 836</td>
</tr>
<tr>
<td></td>
<td>GAM A + C</td>
<td>632</td>
<td>27</td>
<td>68 618</td>
<td>11 180</td>
<td>11 945</td>
<td>11 734</td>
</tr>
<tr>
<td></td>
<td>GAM A + P + C</td>
<td>252</td>
<td>35</td>
<td>32 218</td>
<td>8 711</td>
<td>7 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A * P</td>
<td>146</td>
<td>69</td>
<td>28 471</td>
<td>6 074</td>
<td>5 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bayes A + P + C</td>
<td>242</td>
<td>36</td>
<td>31 459</td>
<td>9 571</td>
<td>1 768</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P</td>
<td>965</td>
<td>17</td>
<td>99 999</td>
<td>13 500</td>
<td>10 211</td>
<td>10 04</td>
</tr>
<tr>
<td></td>
<td>GAM A + C</td>
<td>245</td>
<td>21</td>
<td>28 786</td>
<td>4 031</td>
<td>4 634</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P + C</td>
<td>98</td>
<td>26</td>
<td>15 026</td>
<td>1 600</td>
<td>1 34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A * P</td>
<td>92</td>
<td>47</td>
<td>18 616</td>
<td>7 8</td>
<td>5 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bayes A + P + C</td>
<td>92</td>
<td>30</td>
<td>15 226</td>
<td>1 49</td>
<td>9 74</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P</td>
<td>357</td>
<td>14</td>
<td>35 385</td>
<td>4 821</td>
<td>11 8</td>
<td>1 03</td>
</tr>
<tr>
<td></td>
<td>GAM A + C</td>
<td>152</td>
<td>18</td>
<td>18 888</td>
<td>2 190</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>GAM A + P + C</td>
<td>109</td>
<td>22</td>
<td>15 426</td>
<td>3 1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A * P</td>
<td>80</td>
<td>41</td>
<td>16 116</td>
<td>1 73</td>
<td>1 63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bayes A + P + C</td>
<td>102</td>
<td>28</td>
<td>15 896</td>
<td>6 1</td>
<td>2 45</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A + P</td>
<td>257</td>
<td>13</td>
<td>28 248</td>
<td>3 351</td>
<td>2 75</td>
<td>2 67</td>
</tr>
<tr>
<td></td>
<td>GAM A + C</td>
<td>173</td>
<td>16</td>
<td>20 258</td>
<td>2 321</td>
<td>1 66</td>
<td>1 58</td>
</tr>
<tr>
<td></td>
<td>GAM A + P + C</td>
<td>107</td>
<td>21</td>
<td>14 848</td>
<td>1 01</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAM A * P</td>
<td>89</td>
<td>32</td>
<td>15 215</td>
<td>5 5</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bayes A + P + C</td>
<td>103</td>
<td>23</td>
<td>14 924</td>
<td>9 4</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†Countries are ordered by decreasing population size.
‡All measures of fit are for data aggregated to 5-year periods. Measures include, in order, the plug-in deviance for fitted data, model complexity estimated using $p_D$, the DIC, 10-fold cross-validation (CV$_{10}$), the plug-in deviance for recent projections, and the predictive deviance for recent projections.
confidence intervals (or credible bounds) for Bayesian age-period-cohort models and GAM A * P are shown by 5-year and single-year periods, respectively. The confidence intervals for the means are tight for both models within fitted data. However, the confidence intervals for the Bayesian model are very much wider outside of observed data, becoming moderately uninformative for 5–10 years outside of fitted data.

The GAM A * P estimates were in better agreement with the validation data than the Bayesian estimates for all five countries, except Sweden. The Bayesian age-period-cohort predictions were above observed rates. The GAM estimates tended to be lower than estimates from the Bayes A + P + C model, showing greater curvature. For Sweden, the rates predicted by the GAM A * P model declined more quickly than the observed rates.

Fig. 2. Assessment of recent short-term predictions for lung cancer mortality rates, by country, for females aged 25–84 years.
Fig. 3. Short-term projections for lung cancer mortality rates, by country, for females aged 25–84 years.

For projections outside of observed data, the familiar pattern of GAM estimates being lower to Bayes estimates could be seen (Figure 3). There was good qualitative agreement between both models for the U.S. and the U.K., predicting a continuing decline in age-standardized rates. For the other three countries, the GAM estimates tended to peak and decline, while the Bayes estimates predicted a rise in Swedish rates or for rates being otherwise relatively stable.

7. Discussion

We have used a class of GAMs to estimate lung cancer projections. The models generalized classical age-period and age-cohort models, but avoided modeling issues associated with them and allowed for formal model comparisons. Although the method used only weak a priori assumptions, the confidence intervals
were not large, in marked contrast to recently promoted Bayesian age-period-cohort methods (Bashir and Estève, 2001).

The predictive performance of five flexible rate models for lung cancer mortality were compared for five countries. Based on model comparisons using assessment data, the GAM using bivariate thin-plate smoothing splines generally outperformed the Bayesian age-period-cohort model. In contrast, model choice based on cross-validation was equivocal. This suggests that the models had similar predictive ability within data, however, their predictive performances outside of fitted data were quite different.

Two contributors to the poor performance of the specific Bayesian model are suggested. First, the confidence intervals for the Bayesian model outside of observed data were very wide, indicating poor model specification. This suggests that second-order autoregressive smoothers are poor predictors outside of fitted data. Second, the mean predictions for the Bayesian model did not compare well with the observed assessment data, suggesting that the functional form outside of observed data is inappropriate for these data.

An important limitation of all of these models is the dependence of the projections on the form of the models, where the modeling can be viewed as an exercise in curve-fitting with extrapolation. The validity of any projection will depend on whether the model class will follow the expected changes over time. The thin-plate smoothing splines assume a globally quadratic structure, hence assuming continuing curvature outside of the observed data. In contrast, the Bayesian age-period-cohort model proposed by Berzuini and Clayton (1994) assumes that the separate effects for age, period, and cohort will follow linear trends outside of the observed data. In practice, empirical evidence suggests that period and cohort effects for various cancers tend to lie between linear and quadratic (Ministry of Health, 2002), so that both model classes may perform poorly depending upon the cancer and the population.

The use of the DIC for comparisons of likelihood-based frequentist models with Bayesian models deserves further attention (Spiegelhalter et al., 2002). The DIC provides a ready method for comparison without the requirement to respecify a model for potentially missing data points. In particular, implementation of cross-validation for Bayesian models can be slow. A further methodological implication is to encourage the broader application of the bootstrap to allow the use of predictive measures for model assessment and selection.

The modeling framework can be extended to include a two-dimensional age-cohort model (A * C), which is closely related to the two-dimensional age-period model (A * P) by fitting the same surface using different coordinate sets. It would also be useful to formally assess whether the models are stable when fitted to small numbers of events such as for rare cancers.

The models can be extended to account for overdispersion. This may have little influence on the results for the GAMs, as the means are expected to be similar for Poisson, quasi-Poisson, and negative binomial models, and the nonparametric bootstrap takes account of overdispersed residuals. For the Bayesian model, Knorr-Held and Rainer (2001) suggested the use of a log-normal random effect to account for overdispersion. This would further increase the variance of the Bayesian estimates. For lung cancer data, the degree of overdispersion is only moderate, where quasi-Poisson GAMs gave fitted scale parameters of the order of 1.2.

Although the results were critical of one particular Bayesian age-period-cohort model, there are other Bayesian models that show potential for modeling rate data. One possible approach is to use a Bayesian formulation of thin-plate splines, which has the advantage of providing a framework for model selection (Wood et al., 2002).

A conclusion of public health importance is that the GAMs predict that female lung cancer rates in these countries will be stable or will begin to decline in the future. Given the dependence of current rates on cumulative exposure, these results will be relatively insensitive to recent changes in smoking behavior, however, changes in the cessation rates may increasingly influence predictions over time. Irrespectively, lung cancer will continue to be an important site for cancer death in the future.
ACKNOWLEDGMENTS

Dr. Clements was supported by a postdoctoral fellowship and by the Health Evaluation Research and Outcomes Network Population Health Capacity Building Grant from the National Health and Medical Research Council of Australia. Drs. Clements and Moolgavkar were supported by Cancer Intervention and Surveillance Modeling Network grant U01-CA97415 from the US National Cancer Institute.

REFERENCES


[Received August 10, 2004; first revision January 24, 2005; second revision March 25, 2005; accepted for publication April 25, 2005]