Research Article

Statistics in Medicine

Received 18 December 2012, Accepted 26 May 2014 Published online 3 July 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6248

A Bayesian generalized age–period–cohort power model for cancer projections

Verena Jürgens,[a,b] Silvia Ess,[c] Thomas Cerny[d] and Penelope Vounatsou[a,b*†]

Age–period–cohort (APC) models are the state of art in cancer projections, assessing past and recent trends and extrapolating mortality or incidence data into the future. Nordpred is a well-established software, assuming a Poisson distribution for the counts and a log-link or power-link function with fixed power; however, its predictive performance is poor for sparse data. Bayesian models with log-link function have been applied, but they can lead to extreme estimates. In this paper, we address criticisms of the aforementioned models by providing Bayesian formulations based on a power-link and develop a generalized APC power-link model, which assumes a random rather than fixed power parameter. In addition, a power model with a fixed power parameter of five was formulated in the Bayesian framework. The predictive performance of the new models was evaluated on Swiss lung cancer mortality data using model-based estimates of observed periods. Results indicated that the generalized APC power-link model provides best estimates for male and female lung cancer mortality. The gender-specific models were further applied to project lung cancer mortality in Switzerland during the periods 2009–2013 and 2014–2018. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: cancer projection; Bayesian inference; power model; Nordpred; lung cancer mortality

1. Introduction

Cancer projections estimate the future burden of the disease. They provide important information for health planning and evaluation of intervention effects, for example, screening or changes in diagnostic techniques.

The most common modeling approaches for projecting cancer rates are time series and age–period–cohort (APC) models [1, 2]. Time series models have been applied assuming a Poisson distribution for the disease counts. The models include autoregressive error terms [3] and/or time trends fitted by linear [1], polynomial [3], piecewise linear [4] or spline curves. APC models typically include three components—*age*, corresponding to the age of death, *period*, representing the time period that the death was recorded, and *cohort*, giving the birth cohort of the person that died. Usually these components are stratified into 5-year intervals that are identified by the relation *cohort=period−age*. The dependence between the components is well known as the identification problem; however, it does not affect estimation of the projected rates. Intervals smaller than 5 years have been considered; however, the smaller the interval, the more likely to lead to sparse or zero death counts. Age is assumed to have the highest effect on cancer mortality and incidence rates, whereas period or cohort effects can be neglected in some settings. Short-term projections ($\leqslant 5$ years) are mainly based on age–period models, and long-term ones are based on APC models. However, depending on the data set, cohort effects might also be important for short-term projections. Preliminary analyses are carried out to assess the contribution of cohort effects.

[a] *Department Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstr. 57, CH-Basel, Switzerland*
[b] *University of Basel, CH-Basel, Switzerland*
[c] *Cancer Registry of St. Gallen–Appenzell, CH-St. Gallen, Switzerland*
[d] *Department of Medical Oncology-Hematology, Kantonsspital St. Gallen, Gallen, Switzerland*
**Correspondence to: Penelope Vounatsou, Department Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstr. 57, CH-Basel, Switzerland.*
[†] *E-mail: penelope.vounatsou@unibas.ch*

Models may include a drift parameter to measure the average trend. The drift is often considered constant throughout the periods or reduced by a fixed amount (i.e. 25%) [5].

The APC models are fitted within a generalized linear model framework. Different link functions can be considered to relate the predictors with the mean [6, 7]. Osmond [8] used a Poisson regression model with a log-link function to estimate age-specific and period-specific rates. Møller [9] proposed the power model, using a power rather than a log-link function with a fixed power equal to 5 derived by empirical estimation. Both the Poisson power-link and log-link models are implemented in the R (R Development Core Team, 2011) software package Nordpred developed by the Cancer Registry of Norway. A Bayesian formulation of the Poisson APC model with log-link was suggested by Bray [10], assigning autoregressive priors to the effect of age, period and cohort. Second-order autoregressive processes were appropriate in several settings. Bayesian APC models have been applied frequently to model, for example, lung cancer mortality in West Germany [11], in France [12] or breast cancer mortality in Spain [13].

Several reviews compared different models for cancer projections [2] showing that the Poisson power-link and Bayesian Poisson log-link model outperformed other approaches [1, 14]. The latter model assumes an exponential growth of the predictions. It has been criticized that it might give extreme predictions [2, 15, 16]. The power-link overcomes this problem; however, it may inaccurately estimate low or even zero counts. The Bayesian model formulation smooths age, period and cohort effects obtaining valid estimates for sparse rates or even zero counts [2, 17]. In addition, it can provide uncertainty estimates of rates and assess model performance taking into account the estimation error. To our knowledge, a Bayesian formulation of the Poisson power-link model has not been developed yet. Furthermore, it is unclear whether all applications of the power-link model support a fixed power parameter equal to 5.

In this paper, we formulated the power model within the Bayesian framework and developed a generalized model considering a random rather than fixed power parameter. The aforementioned models were compared with the ones using the log-link function. All models were applied on gender-specific lung cancer mortality data from Switzerland and compared with the fixed power-link and log-link model fitted in Nordpred using maximum likelihood. The sum of squared residuals (SSR) of the projected rates was calculated for all models, while the logarithmic score (LS) [18] and deviance information criterion (DIC) was applied to the Bayesian models to assess their predictive performance. The gender-specific model with the best accuracy was employed to project lung cancer mortality rates for the periods 2009–2013 and 2014–2018.

2. Methodology

2.1. Model formulations

In this study, we developed Bayesian APC models assuming that the observed age-specific and period-specific death counts N_{ij} follow a Poisson $N_{ij} \sim \text{Pois}(\mu_{ij})$ distribution and using a model-specific link-function g.

The mean of the age-specific and period-specific death counts μ_{ij} was regressed on the effects of age, period and cohort, using the corresponding population as the offset. In particular, let α_i, β_j and γ_k $(k = I - i + j)$ be the effects of age, period and cohort, and N_{ij} and M_{ij} be the age-specific and period-specific death counts and population, respectively.

$$g\left(\frac{\mu_{ij}}{M_{ij}}\right) = \alpha_i + \beta_j + \gamma_k \tag{1}$$

Under a log-link as well as power-link function, the regression model relating the mortality rate to the time effects reads as follows:

$$\begin{aligned} \textit{Log-link model}: &\quad \log(\mu_{ij}) = \log(M_{ij}) + \alpha_i + \beta_j + \gamma_k \\ \textit{Power-link model}: &\quad \mu_{ij}^{\frac{1}{5}} = M_{ij}^{\frac{1}{5}}\left(\alpha_i + \beta_j + \gamma_k\right) \end{aligned} \tag{2}$$

In the aforementioned model, the power is assumed to be fixed. We extended the model by considering the power to be a parameter estimated by the data and called it generalized APC power model (GPM). The power w was assumed to be random, as the degree of the growth of the rate might change, depending

on the given data set. The model was formulated as follows:

$$GPM: \qquad \mu_{ij}^{\frac{1}{w}} = M_{ij}^{\frac{1}{w}} \left(\alpha_i + \beta_j + \gamma_k\right) \tag{3}$$

The aforementioned model overcomes estimation of extreme rates that may be obtained from the log-link. Bayesian simulation-based estimation allows flexible prior specification that smooths time effects.

2.2. *Prior specifications*

To complete Bayesian specification, prior distributions were assigned to the parameters. Following Bray *et al.* [10], time effect-specific (i.e. age, period and cohort) smoothing prior formulations were considered. The first two groups of each time effect were restricted to follow a normal distribution with mean 0 and vague precision. The remaining groups were assumed to follow a second-order autoregressive process, depending each effect on its two predecessors, to allow for dependence and smoothing [10]. The aforementioned prior distributions are formulated as follows (illustrated for the effect of age):

$$\alpha_1 \sim \mathcal{N}\left(0, 1.0E{-}06 * \tau_\alpha\right) \tag{4}$$

$$\alpha_2 | \alpha_1 \sim \mathcal{N}\left(0, 1.0E{-}06 * \tau_\alpha\right) \tag{5}$$

$$\alpha_i | \alpha_{1,\ldots,i-1} \sim \mathcal{N}\left(2\alpha_{i-1} - \alpha_{i-2}, \tau_\alpha\right), \text{for } 3 \leqslant i \leqslant I \tag{6}$$

where I is the total number of age groups. Non-informative priors were assigned to the precision parameters τ_α, τ_β and τ_γ for the effect of age α, period β and cohort γ, respectively. For comparison purposes, we used the same prior distributions across models.

A discrete prior distribution was assigned on the power parameter w, assuming that $w \sim cat(p[])$. This prior formulation specifies a set of possible values $w \in \{1, \ldots, n\}$ with probabilities $p[w = i] = p_i$ and $\sum_{i=1}^{n} p[i] = 1$. In our specification, we consider that w could take any integer from 1 to 10, each with probability equal to 0.1.

3. Assessing model predictive performance

Empirical projections, estimating mortality for observed periods, have been carried out applying the different models, that is, Bayesian log-link, power-link model and the GPM, the Nordpred log-link and power-link models. In addition, the performances of the log-linear model (excluding the cohort effect), which is generally used for short-term projections [19], and the simple averaging method [20] have been assessed. The latter has been recommended for the projection of low count data and is based on the assumption that the population size will change in the future, but the mortality rate remains the same as in the current period. Results have been validated by the LS, DIC, SSR of each model m as well as graphical assessment of the projected rates.

$$\mathrm{SSR}(m) = \sum_{i,j} \frac{\left(\mathrm{R_{ij}} - \hat{\mathrm{R}}_{\mathrm{ij}}\right)^2}{\mathrm{R_{ij}}} \tag{7}$$

where $\mathrm{R_{ij}}$ and $\hat{\mathrm{R}}_{\mathrm{ij}}$ are the observed and estimated rates for age group i and period j, respectively. The LS was calculated following Ntzoufras [18] as follows:

$$\mathrm{LS}(m) = -\sum_{i,j} \log \mathrm{PPO}_{ij}(m) \tag{8}$$

where PPO is the posterior predictive ordinate.

4. Application

In Switzerland, the National Program Tobacco 2008–2012 (NPT 2008–2012) was launched to reduce tobacco-related morbidity and mortality in the country. The initiative strategic goals were set until the end of 2012 to accomplish the mission. Projections of tobacco-related cancer mortality based on past trends, not taking into account recent interventions as the NPT 2008–2012, may increase awareness on the program. However, interpretation on the impact of the program based on the projected rates should be carried out carefully as the difference between estimated and observed might be due to several factors, and interventions are only one possibility among others.

Count mortality data from death certificates and population size at national level were provided from the Swiss Federal Statistical Office (FSO). The time range considered for both data sets was 1974–2008 and was split into seven periods (1974–1978,..., 2004–2008). Age was classified in 5-year groups (25–29,..., older than 85 years), and 19 ten-year overlapping cohorts were constructed following the relation cohort = period − age. Gender-specific mortality rates were age-standardized directly using Segi World population [21]. In 1995, the rules related to coding death certificates have been changed in Switzerland—priority of certain causes has been removed from the regulations. Therefore, the number of these cause-specific deaths reported before and after this changes differed and have to be adjusted. The Swiss Federal Statistical Office provides a disease-related correction factor to adjust death reports before 1995 [22].

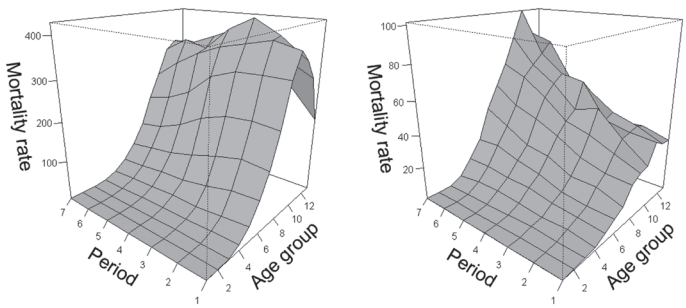
Exploratory analysis of tobacco-related cancer mortality from 1974 to 2008 indicated different trends for each gender (Figure 1). Rates for men remain stable since the last decade, while female lung cancer mortality increased steadily.

Empirical projections were carried out for the two last periods (1999–2003, 2004–2008) with known mortality rates to validate the model performance. After validation, the model with the best fit was applied to project gender-specific lung cancer mortality in Switzerland for the periods 2009–2013 and 2014–2018. In addition, predictive performance of the models was assessed for a shorter period of observed mortality, that is, 1984–1988,..., 2004–2008. Model formulation and fitting was based on Markov chain Monte Carlo (MCMC) [23] and was performed in WinBUGS (Imperial College and MRC, London, UK). Convergence was assessed by the Gelman and Rubin diagnostic test within the R package coda.

4.1. Model predictive performance

Exploratory analysis assuming different drifts indicated no effect on the fitting and the estimates. Therefore, the drift parameter was considered to be zero. Model predictive ability was evaluated by different metrics. Results were compared with the Nordpred equivalent whenever possible (Table I). Firstly, the SSR was calculated for all models, and visual assessment was carried out by comparing model-based estimates with the observed rates for the projected periods. As a next step, Bayesian models were compared on the basis of the LS and DIC.

The results of the empirical projections are shown in Table I. SSR and visual assessment of the projected rates (Figure 2) indicated that Bayesian models performed better than the equivalent implemented in Nordpred. For men, the power-link outperformed the log-link model. The latter provided extremely



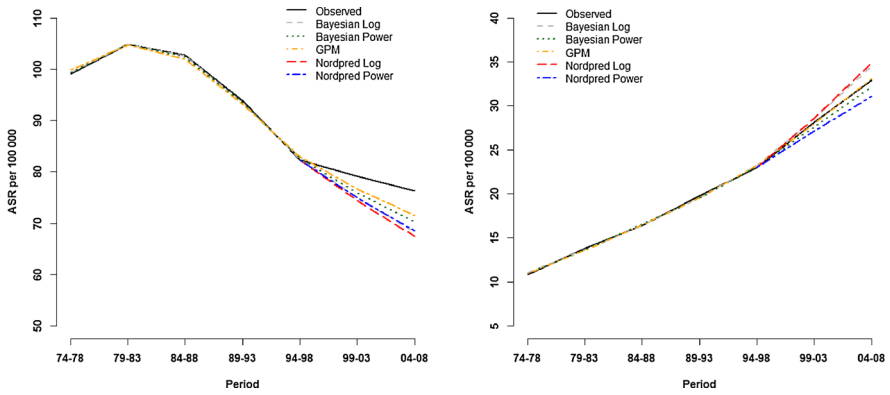
(a) (b)

Figure 1. Observed crude lung cancer mortality rate (per 100 000 inhabitants) for men (a) and women (b) from 1974–1978 to 2004–2008.

Table I. Sum of squared residuals (SSR) of empirical projections of gender-specific lung cancer mortality rates, logarithmic score (LS) and deviance information criterion (DIC) for the Bayesian models.

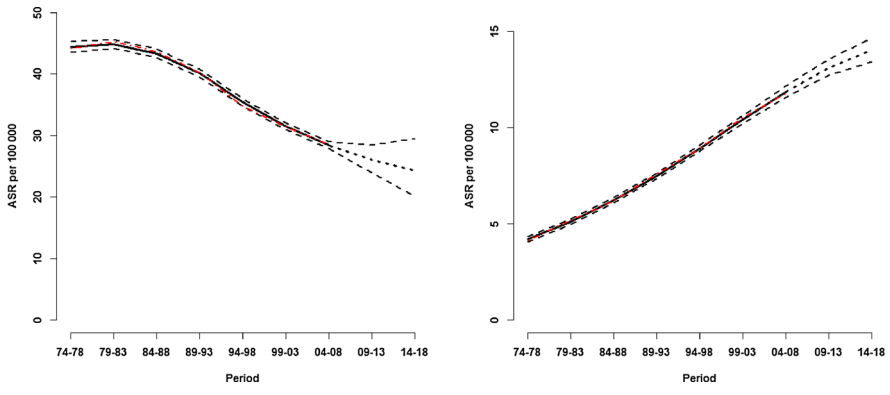
	Male			Female		
	LS	DIC	$SSR_{6,7}$	LS	DIC	$SSR_{6,7}$
Bayesian						
Log-link	−207.7	589.2	40.1	−207.4	494.3	8.9
Power-link	−206.5	590.3	30.9	−208.4	490.4	9.9
Log-linear	−340.5	868.4	142.6	−259.5	516.6	13.1
GPM	−205.7	597.3	21.9	−206.5	496.8	8.6
Nordpred						
Log-link			53.1			13.7
Power-link			41.6			11.4
Simple averaging			60.1			86.4

GPM, generalized age–period–cohort power-link model.



(a) (b)

Figure 2. Observed and predicted Swiss lung cancer mortality rates for men (a) and women (b).



(a) (b)

Figure 3. Observed (red), model-based estimates of fitted (1974–1978 to 2004–2008) and projected (2009–2013, 2014–2018) Swiss age-standardized lung cancer mortality rate for men (a) and women (b) (per 100 000 inhabitants) with 95% credible bounds (CB).

low estimates. For women, the log-link was better than the power-link; however, it overestimated mortality. The Bayesian log-linear model and the simple averaging method performed worst. Best predictive ability for both genders was given by the GPM, as indicated by the SSR, LS and the plotted estimates of

Table II. Model-based estimates (GPM) of fitted and projected age-standardized gender-specific lung cancer mortality rate (per 100 000 inhabitants) with 95% CB in brackets.

	Fitted rates							Projected rates	
	74–78	79–83	84–88	89–93	94–98	99–03	04–08	09–13	14–18
Male									
Overall	44.35 (43.52;45.21)	44.84 (44.14;45.55)	43.32 (42.62;44.01)	40.09 (39.50;40.78)	35.38 (34.75;35.94)	31.49 (30.95;32.02)	28.41 (27.87;28.98)	26.08 (23.93;28.46)	24.28 (20.17;29.55)
25–44 years	1.07 (0.98;1.17)	1.00 (0.93;1.09)	0.98 (0.91;1.06)	0.88 (0.82;0.95)	0.72 (0.65;0.78)	0.62 (0.56;0.68)	0.57 (0.50;0.65)	0.57 (0.41;0.80)	0.59 (0.31;1.19)
45–64 years	20.64 (20.11;21.16)	19.94 (19.52;20.38)	18.40 (17.99;18.80)	16.02 (15.68;16.41)	13.55 (13.20;13.86)	12.09 (11.79;12.40)	11.20 (10.89;11.53)	10.41 (9.37;11.56)	9.79 (7.89;12.26)
65–84 years	21.65 (21.22;22.09)	22.71 (22.33;23.10)	22.62 (22.25;22.97)	21.75 (21.40;22.11)	19.69 (19.37;20.02)	17.46 (17.16;17.75)	15.41 (15.11;15.72)	13.92 (13.00;14.95)	12.89 (11.11;15.04)
Older 85 years	1.00 (0.89;1.11)	1.18 (1.12;1.26)	1.33 (1.27;1.39)	1.44 (1.38;1.50)	1.42 (1.36;1.47)	1.32 (1.26;1.37)	1.23 (1.18;1.28)	1.15 (1.06;1.24)	1.00 (0.85;1.18)
Female									
Overall	4.21 (4.07;4.36)	5.12 (4.98;5.26)	6.22 (6.08;6.36)	7.50 (7.35;7.63)	8.94 (8.78;9.09)	10.44 (10.23;10.63)	11.87 (11.59;12.17)	13.12 (12.72;13.57)	14.03 (13.42;14.64)
25–44 years	0.27 (0.24;0.30)	0.35 (0.32;0.39)	0.45 (0.41;0.50)	0.52 (0.48;0.57)	0.52 (0.48;0.57)	0.47 (0.42;0.51)	0.39 (0.33;0.45)	0.34 (0.26;0.44)	0.31 (0.19;0.49)
45–64 years	1.96 (1.87;2.06)	2.40 (2.30;2.50)	2.97 (2.85;3.08)	3.67 (3.56;3.79)	4.46 (4.32;4.58)	5.25 (5.09;5.40)	5.92 (5.71;6.14)	6.21 (5.91;6.54)	6.01 (5.60;6.44)
65–84 years	1.82 (1.75;1.91)	2.20 (2.13;2.28)	2.62 (2.54;2.70)	3.11 (3.03;3.19)	3.72 (3.63;3.82)	4.42 (4.31;4.55)	5.22 (5.07;5.37)	6.17 (5.96;6.40)	7.22 (6.93;7.54)
Older 85 years	0.15 (0.13;0.18)	0.16 (0.15;0.18)	0.18 (0.16;0.19)	0.20 (0.18;0.22)	0.24 (0.23;0.26)	0.30 (0.28;0.32)	0.34 (0.32;0.36)	0.40 (0.38;0.43)	0.48 (0.45;0.51)

GPM, generalized age–period–cohort power-link model; CB, credible bounds.

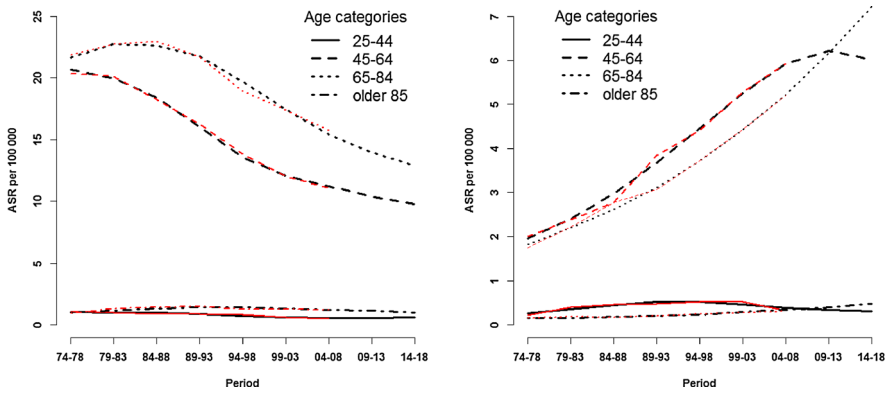


Figure 4. Observed (red), model-based estimates of fitted (1974–1978 to 2004–2008) and projected (2009–2013, 2014–2018) Swiss age-standardized lung cancer mortality rate for men (a) and women (b) (per 100 000 inhabitants) by age category.

the empirical projection. The posterior median of the random power parameter for men and women was estimated to be 3 (3;3) and 7 (7;7), respectively. For men, the mortality rates level off during the projected period. Female mortality rates continue increasing during 1999–2008 following a slightly steeper trend than during 1974–1998.

Empirical projections based on a shorter period of observed mortality (1984–1988,..., 2004–2008) indicated best performance of the Bayesian log-link model for men. However, the predictive performance of the GPM based on the longer observed period was better than the one based on the shorter fitting period. For women, the GPM indicated best predictive ability for the shorter as well as for the longer fitting period; however, the latter provided better fit.

4.2. Projections

Gender-specific lung cancer mortality rates were projected for the periods 2009–2013 and 2014–2018, on the basis of the observed data from 1974 to 2008. Future rates were predicted for each gender by using the GPM and the Bayesian log-link model for men and women, respectively. On the basis of evaluation, different models were applied for gender-specific projections. To account for potential age-specific trends, age was split into four age categories 25–44, 45–64, 65–84, and 85 years and older. Overall and age category-specific rates were age-standardized by Segi World population. Observed lung cancer mortality rates by gender are illustrated in Figure 1. Figure 3 and Table II indicate a steadily declining trend for male lung cancer mortality since 1979–1983. This can be seen for all groups, except the ones older than 85 years, which shows a rising trend until 1993, followed by a decline (Figure 4).

Female lung cancer mortality was estimated to be more than three times higher in 2014–2018 in comparison with 1974–1978. Age category-specific plots present a stable trend for women aged 45–64 years in 2004–2008.

Final projections for men and women estimated the random power parameter to be 3 and 7, respectively.

5. Discussion

In this paper, we developed a Bayesian APC power model with fixed power of 5 and a generalized APC power model for cancer mortality/incidence projections, which overcomes limitations of well-established models. The commonly used power model was formulated within the Bayesian framework, and it was further extended to allow for a random power parameter instead of the fixed one proposed by Møller [9].

Model performance was compared with the frequently used Bayesian APC model with a log-link function. We assumed that the counts arise from a Poisson distribution. The models were compared with the commonly used ones implemented in the Nordpred software assuming the log-link and the power(5)-link. The predictive performance of the new models was evaluated on Swiss lung cancer mortality data using model-based projections of observed periods. The model with the best fit was applied to project

age-specific and gender-specific lung cancer mortality rates in Switzerland for the periods 2009–2013 and 2014–2018.

The GPM performed best for men and women. For both genders, the log-linear and simple averaging method gave the poorest fit. The latter assumes continuation of the trend observed in the recent period and is sensitive to the projected population [20]. For the empirical projection, the population was known, which suggests that the assumption of a continuing trend was violated. Figure 2 illustrates the changing trend for both genders—a flattening decrease for men and a steepening increasing trend for women.

The unsatisfactory performance of the log-linear model can be explained by the missing cohort effect, which was found to be strongly related to lung cancer mortality rates [24]. This relation can be explained by the gender-specific tobacco epidemics.

Møller *et al.* [25] compared several projection methods, for example, APC models using a power-link or a log-link on several cancer incidence data from five Nordic countries. They also found that models using a power instead of a log-link function have better predictive ability. The authors assessed models with power parameters fixed to 5 and 2 and concluded that for short-term predictions ($\leqslant$5 years), the model based on the power of 2 outperformed the one with a power of 5. This result was not seen for the long-term projections, and they pointed out that too small values may implicate numerical problems.

The Canadian research group from the Alberta Health service evaluated long-term projections of cancer mortality and incidence data based on Nordpred, generalized additive and Bayesian models [2]. The Bayesian specifications included the log-link assuming different prior distributions for the effects of age, period and cohort. They concluded that the models in Nordpred using power 5 and the Bayesian model using second-order autoregressive priors for the age, period and cohort effects performed best; however, the Bayesian one performed better considering mortality projections, and the Nordpred achieved better results for incidence. Mortality is more sparse than incidence and therefore more challenging for the modeling. The latter is true for cancer sites with high survival rates. Some authors [2] argue against the use of Bayesian APC models because of the complexity of the MCMC fitting; however, over the last years, the number of scientific publications, using the Bayesian framework, is increasing steadily.

The analysis of the Swiss lung cancer mortality data indicated an ongoing decline for overall male mortality from 1974 onwards. Age-specific mortality followed the overall trend with exception the group of 65–84 years old, whose decreasing trend started after 1979. Despite that mortality rates are decreasing, the number of projected deaths increased after 2009. Female mortality increases since 1974 for all age groups; however, the group of 45–64 years old reaches a plateau around 2004–2008.

In general, results of projections should be interpreted carefully, as they depend on different factors such as model choice. Clements *et al.* [26] analyzed lung cancer mortality for women in five countries. They compared different approaches, which provided different trends for the future rates. The authors pointed out that the dependence of the projections on the form of the model is an important limitation. Furthermore, cancer projections are sensitive to population projections because the proportion of elderly people influences mortality rates and crude death counts. In our analyses, population data for 2009–2013 and 2014–2018 were provided by the Swiss Federal Office of Statistics, which extrapolates populations under different scenarios regarding migration, birth and death rates. Our population estimates are based on non-extreme demographic changes.

Projections of lung cancer mortality are based on past trends, not taking into account recent interventions. The NPT 2008–2012 was launched to reduce tobacco-related cancer morbidity and mortality in the country. The initiative strategic goals were set until the end of 2012. Our findings could help to evaluate the program, as model-based estimates can be compared with observed data in the future.

Several metrics exist to assess the predictive performance of a model [27]. We used the SSR [17] as well as the DIC and the LS [18] for the Bayesian models. In addition, the estimated rates have been plotted, and predictive ability has been assessed graphically. For men and women, the different metrics gave the same conclusion.

In this paper, different functions have been applied to link the expected value of the mortality to the predictor. Their choice depend on the data set under consideration. While a log-link function is more suitable to model a steep trend, a power-link function allows the specification of a certain degree of increase or decrease of the trend. In this study, gender-specific trends have been observed. Men had a rather decreasing trend that flattened at the end of the study period. For women, an increasing trend was observed. Models based on the log-link function provided the most extreme estimates, underestimating and overestimating the mortality rate for men and women, respectively (Figure 2). Estimates of the power-link models were less extreme and performed better, while the GPM performed best in both settings. Visualization of the estimated rates by the different link-functions clearly shows the underestimation and

overestimation by the log-link and power-link models and indicates that a power parameter lower/greater than 5 would be more appropriate to model Swiss male/female lung cancer mortality. This was confirmed by the estimated power of 3 and 7 by the GPM for men and women, respectively.

The proposed models extend the list of APC models, and as already discussed earlier, they had better predictive performance in modeling the lung cancer data of our application. The models have a number of strengths that make them especially suitable for modeling low counts as we have shown in the analysis of female mortality data. Lee *et al.* [1] and Qiu *et al.* [20] suggested that the average counts in the few years prior to projection may be a reasonable approach for sparse or low count data. Our analysis showed that the simple average method performed worse than any other approach in the case of female low count mortality data and that Bayesian formulations outperform non-Bayesian approaches especially when modeling low counts. We are currently assessing the models in projecting cause-specific mortality other than lung cancer.

The Bayesian model formulations we developed can be extended to account for extra Poisson variation, present in sparse cancer data, by considering a negative binomial distribution that is a mixture of a Poisson distribution with a Gamma mean parameter [18]. However, no over-dispersion or under-dispersion was present in the data set under consideration. In fact, the negative binomial model did not improve the predictive performance of the Poisson model. Choosing a distribution other than the Poisson might be reasonable for the analysis of low count data, for example, at sub-regional level, which would allow the formulation of a subgroup-specific dispersion parameter. Furthermore, potential heterogeneity can be captured by introducing random effects; however, this specification will increase the number of parameters to be estimated.

Our analyses projected mortality rates at national level. Regional differences can be important for health planning purposes. Model formulations can be easily extended to assess regional differences and spatio-temporal patterns across the country.

Acknowledgements

This work was supported by Oncosuisse (grant number KLS-02393-02-2009) and the Swiss National Science Foundation (grant number 32003B_135769). We would like to thank the Swiss Federal Statistical Office for providing the data.

We gratefully acknowledge the critical reviews and helpful suggestions by the anonymous reviewers the paper benefited from.

References

1. Lee TCK, Dean CB, Semenciw R. Short-term cancer mortality projections: a comparative study of prediction methods. *Statistics in Medicine* 2011; **30**(29):3387–3402.
2. Qiu Z, Jiang Z, Wang M, Hatcher J. Cancer Projections Network (C-Projections). Long-term projection methods: comparison of age-period-cohort model-based approaches, 2010.
3. Wingo PA, Landis S, Parker S, Bolden S, Heath CW. Using cancer registry and vital statistics data to estimate the number of new cancer cases and deaths in the United States for the upcoming year. *Journal of Registry Management* 1998; **25**(2): 43–51.
4. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* 2000; **19**:335–351.
5. Møller B, Fekjaer H, Hakulinen T, Tryggvadttir L, Storm HH, Talbäck M, Haldorsen T. Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention* 2002; **11 Suppl 1**:S1–96.
6. Bani KM, Gelfand AE. Generalized linear models with unknown link functions. *Biometrika* 1994; **81**(2):237–245.
7. Pregibon D. Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 1980; **29**(1):15–24.
8. Osmond C. Using age, period and cohort models to estimate future mortality rates. *International Journal of Epidemiology* 1985; **14**(1):124–129.
9. Møller B. Prediction of cancer incidence – methodological considerations and trends in the Nordic countries 1958–2022. *PhD Thesis*, Faculty of Medicine, University of Oslo, Unipuc AS, Oslo, 2004.
10. Bray I. Application of Markov chain Monte Carlo methods to projecting incidence and mortality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2002; **51**(2):151–164.
11. Knorr-Held L, Rainer E. Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. *Biostatistics* 2001; **2**(1):109–129.
12. Eilstein D, Zoé U, Lim TA, Bloch J. Lung cancer mortality in France. Trend analysis and projection between 1975 and 2012, using a Bayesian age-period-cohort model. *Science Direct* 2008; **59**(3):282–290.
13. Clèries R, Ribes J, Esteban L, Martinez JM, Borrs JM. Time trends of breast cancer mortality in Spain during the period 1977-2001 and Bayesian approach for projections during 2002-2016. *Annals of Oncology* 2006; **17**(12):1783–1791.

14. Cancer Projections Network (C-Projections). Comparison of short-term projection methods: proposing the hybrid approach and Bayesian models, Alberta Health Services, 2010.
15. Bray F, Møller B. Predicting the future burden of cancer. *Nature Reviews Cancer* 2006; **6**(1):63–74.
16. Møller B, Weedon-Fekjaer H, Haldorsen T. Empirical evaluation of prediction intervals for cancer incidence. *BMC Medical Research Methodology* 2005; **5**(21).
17. Baker A, Bray I. Bayesian projections: what are the effects of excluding data from younger age groups? *American Journal of Epidemiology* 2005; **162**(8):798–805.
18. Ntzoufras I. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2009.
19. Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine* 1994; **13**:1513–1523.
20. Qiu Z, Hatcher J, Wang M. Cancer Projection Analytical Network Working Team. Review Cancer Projections Methods for Canadian Partnership Against Cancer Analytical Network, *Technical Report*, Alberta Health Services.
21. Segi M. *Cancer mortality for selected sites in 24 countries (1950–57)*, 1960.
22. Berrut S, Junker C. Von Generation zu Generation: Entwicklung der Todesursachen 1970 bis 2004, Bundesamt für Statistik, Neuchâtel, Schweiz, 2008.
23. Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, UK, 1996.
24. Lee PN, Fry JS, Forey BA. Trends in lung cancer, chronic obstructive lung disease, and emphysema death rates for England and Wales 1941–85 and their relation to trends in cigarette smoking. *Thorax* 1990; **45**:657–665.
25. Møller B, Fekjaer H, Hakulinen T, Sigvaldason H, Storm HH, Talbäck M, Haldorsen T. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Statistics in Medicine* 2003; **22**(17):2751–2766.
26. Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics* 2005; **6**(4):576–589.
27. Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Biometrics* 2009; **65**(4):1254–1261.