Using the age-period-cohort model to describe and predict mortality from major smoking-related diseases

Peter N Lee, John Fry and Jan Hamling

From P N Lee Statistics and Computing Ltd, 17 Cedar Road, Sutton, Surrey, SM2 5DA, United Kingdom

Summary

Introduction and objectives

The age-period-cohort (APC) model is a method for fitting age- and period-specific mortality rates, typically based on death and population data for a period which are presented in five-year groups. The model fits rates as a product of age, period (at death) and (birth) cohort effects. The fitted model can also be used to predict rates for a future period, using various alternative methods for extrapolating period and cohort effects. The objectives of the work described here are: (i) to investigate the adequacy of APC model fit and predictions based on sex-specific data for four major smoking-related diseases (lung cancer, chronic obstructive pulmonary disease [COPD], ischaemic heart disease [IHD], and stroke) and for 12 developed countries, (ii) to predict future rates, and (iii) to investigate how well fitted birth cohort effects correlate with cohort-specific cigarette consumption estimates.

Methods

Separately for each combination of sex, disease and country, the APC model was first used to summarize the 1961-1990 rates and to study goodness-of-fit. The model fitted to the 1961-1990 data, in conjunction with various alternative approaches for extrapolating the fitted period and cohort effects, was then used to predict rates for 1991-2010, and to study goodness-of-fit of the predictions to the known data. Based on this work a "best" extrapolation approach was then selected. Next, the APC model was fitted to 1981-2010 rates, and goodness-of-fit was studied. The model fitted to the 1981-2010 data was then used, both to predict the unknown rates for 2011-2030 using the "best" extrapolation approach, and to compare the fitted cohort effects to corresponding estimates of cigarette consumption per head.

Results

The model explains much of the variation in rates unexplained by age, with cohort typically explaining more than period. The fit of predicted to known rates is poorer, though predictions using the model are in most cases, though not always, much better than simpler predictions which ignore the period and cohort effects. Some misfits arise where the observed 1991-2010 rates could not reasonably have been predicted from 1961-1990 data by any model which ignores trends in risk factors. For lung cancer, particularly, but not for IHD and stroke in females, cohort values are consistently associated with cigarette consumption. With

exceptions, predicted 2026-2030 rates, particularly for IHD and stroke, are much lower than in 1981-85.

Conclusions

The model seems useful for summarizing rates and (within the context of a model incorporating no risk factor data) for predicting future rates.

Contents

1 Introduction

Being able to predict mortality rates from smoking-related diseases is often useful. Thus, one may wish to compare future deaths attributable to smoking-related diseases assuming that existing trends in smoking habits remain unchanged, with those assuming that some smokers switch to a reduced risk product. Predicting mortality rates accurately is difficult, particularly where future changes in factors causing, preventing or curing a disease, or in diagnostic standards, cannot reliably be known. However, it may be useful to generate predictions assuming that existing mortality trends continue. The objective of the work described here is to investigate the usefulness of a relatively simple approach which fits a model to existing rates, then uses it to generate predictions. The model selected involves no risk factors, such as smoking, relying instead on the existing pattern of mortality rates to reflect underlying risk factor trends.

The model we investighated is the Osmond and Gardner age-period-cohort (APC) model [1, 2], which analyses mortality data (deaths and person-years at risk) portrayed in a matrix with five-year age groups as rows, and five-year time periods as columns. Presented thus, diagonals of increasing age and period relate to "cohorts" born in a ten year period around a specific year. Mortality rates are modelled as the product of age, period and cohort values, a problem with no unique solution, since knowledge of any two of age, period (of death) and (birth) cohort defines the other. Osmond and Gardner [2] approached this problem by first solving each two-parameter submodel (age-period, age-cohort and period-cohort) and then minimizing a function of the three weighted squared differences between each two-factor submodel and the full three-factor model, the resulting solution giving more weight to the two-parameter models giving better fits to the data. Their model (or a Bayesian form of it) has been widely used to describe existing and predict future mortality rates for various diseases (e.g. [3-13]).

The set of age, period and cohort values generated by fitting the model to the known data are not fully sufficient to derive predictions, as extrapolated period values are required for the projected periods, while extrapolated cohort values are required for later-born cohorts not contributing to the dataset (though these relate only to the younger age groups). Different extrapolation approaches were used in the papers referred to above, one of our objectives being to choose from some of the possible approaches.

Here, we examine the APC method in detail using data for both sexes and four major smoking-related diseases (lung cancer, COPD, IHD and stroke) from 12 developed countries for a 50-year period.

We first study model fit to the 1961-1990 data, and then compare predictions, based on the fitted model parameters and using alternative cohort and period extrapolation approaches, to the known rates observed in 1991-2010. We then fit the model to 1981-2010 data, deriving projected rates for 2011-2030 using the extrapolation approach which was found to work best when predicting the known rate for 1991-2010. Finally, we compare fitted cohort values with cohort-specific cigarette consumption estimates.

2 Methods

2.1 Data

Sex-specific data were obtained for 10 five-year age groups (30-34, … 75-79) and 10 five-year periods (1961-1965, … 2006-2010) for the four smoking-related diseases and for 12 developed countries (Austria, Canada, France, Germany, Hungary, Italy, Japan, Poland, Sweden, Switzerland, UK and USA). Exceptionally, data for Germany were only available from 1981.

Data on numbers of deaths came from the WHO [14]. The ICD codes used were lung cancer (revision 7: A050, 8: A051, 9: B101, 10: C33, C34), COPD (7: A093, A097, 8: A093, A096, 9: B323-B325, 10: J40-J47, J67), IHD (7: A081, 8: A083, 9: B27, 10: I20-I22, I24, I25) and stroke (7: A070, 8: A085, 9: B29, 10: I60-I64, I67, I69). Exceptionally, Switzerland used special codings for ICD revisions 9 and 10 (lung cancer 1034, COPD 1076, IHD 1067, stroke 1069). The time of switching between revisions varied between country (see Table 1).

Missing data (see also Table 1 footnotes) were replaced by linear extrapolation for Canada and France (using data for 2005-2009), and by linear interpolation for Italy (using data for 1999-2003 and 2006-2010), and Poland (using data for 1992-1996 and 1999-2003). Zero deaths in one case (COPD, Switzerland, 2006-2010, age 30-34) were replaced by 0.1, to allow the use of logarithms.

UN population data [15] were used, the data for 2011-2030 being the UN's "medium fertility variant projection".

WHO population data [14] were also available, but were not used for various reasons. Firstly, WHO provides no projected estimates. Second, there are some missing data for 2005-2010 for WHO but not UN. Third, preliminary analyses predicting 1991-2010 from 1961-1990 rates produced no better predictions on average using WHO data (results not shown).

Data for age groups above 75-79 were not used, due to grouping into 80+ years for some countries. Data for age groups below 30-34 were not used due to small numbers of deaths, particularly for lung cancer and COPD.

2.2 Fitting the APC model

The APC model was separately fitted for 1961-1990 and 1981-2010 for each sex/disease/country combination (except for Germany for 1961-1990), using the Osmond and Gardner [2] method. The model was fitted to a set of rates, r_{ij}, where i is age group (i=1 to 10) and j period (j=1 to 6). As shown in Figure 1, which relates to predicting data for 2011-2030 from known data for 1981-2010, the subscript k (=j–i+10) relates to successive cohorts (k=1 to 15), with k=1 representing the oldest age group (10) in period 1 and k=15 the youngest (1) in period 6. The following models were fitted by a general least-squares approach, weighted on d_{ij} , the number of deaths for age i and period j.

age model (A)	$\log r_{ij}$	=	$\log a_i$
age-period model (AP)	$\log r_{ij}$	=	$\log a_i + \log b_j$
age-cohort model (AC)	$\log r_{ij}$	=	$\log a_i + \log c_k$
period-cohort model (PC)	$\log r_{ij}$	=	$\log b_j + \log c_k$
age-period-cohort model (APC)	$\log r_{ij}$	=	$\log a_i + \log b_j + \log c_k$

Thus, for the APC model, the following function was minimized:

$$f(\boldsymbol{a}, \boldsymbol{p}, \boldsymbol{c}) = \sum_{i,j} d_{ij}\left(\log r_{ij} - log\ a_i - log\ b_j - log\ c_{1-i+j}\right)^2$$

Period values (b_j) and cohort values (c_k) were given an "average" value of unity, making the age values (a_i) similar to the age-specific death rates. While the best-fitting parameter values for the A, AP, AC and PC models were derived by standard methods, those for the APC model were derived using the Osmond and Gardner [2] method. Fuller details of the method and the definition of the constraints on the period and cohort values are given in the source paper [2].

For each model, not only were the parameter values estimated, but also the residual sum of squares (RSS) and the mean residual sum of squares (MRSS), based on the degrees of freedom (DF) for the various models (A 51, AP 45, AC 36, PC 40, APC 32). These were then used to estimate various percentages of variance explained, e.g. the percentage of the RSS from A explained by APC (expected to be 100(51-32)/51 = 37.3% without period or cohort effects), or the percentage of the RSS from AP explained by APC (expected to be 100(45-32)/45 = 28.9% without cohort effects). Approximate goodness-of-fit chi-squared statistics were also estimated by summing $X_{ij} = (O_{ij} – E_{ij})^2 / E_{ij}$ over each cell, where O_{ij} is the observed number of deaths, and E_{ij} is that expected by multiplying the fitted rate by the population. Plots of observed and fitted rates, and inspection of X_{ij} values for individual cells were also used to investigate misfit.

2.3 Extrapolating cohort and period effects

Unpublished work which we conducted in the 1980s using the APC model limited the approaches pursued here. Thus, linear extrapolation of cohort and period values, possibly producing negative predictions, was not studied, attention being restricted to log-linear extrapolation. The earlier work also recommended using weights decreasing exponentially in the past, to allow recent values to have greater influence. Where cohort or period values show a well-defined peak (or trough), predictions might theoretically be improved by basing extrapolations only on post-peak values. However, the peak was often poorly defined or non-existent so this approach was not pursued.

Three approaches were used to extrapolate cohort values. In Approach 1, values for cohorts not contributing to the original data (cohorts 16 to 19 – see Figure 1) were estimated by weighted linear regression of log cohort values for the known cohorts. The weights used were

powers of two, so the earliest cohort had weight = 1 and the most recent weight = 16284. As the fitted values for the most recent cohorts are based on limited data, often with few deaths, the alternative approaches used only the more reliably based cohort values. Thus, in Approach 2, the final two fitted cohort values (c_{14} and c_{15}) were ignored, these values (and also those for the later cohorts) being estimated by weighted extrapolation. Approach 3 was similar, but here only c_{15} was ignored.

Two approaches were used to extrapolate period values. In Approach A, period values were extrapolated based on the ratio of the last two values, b_5 and b_6. Thus, if $U = b_6/b_5$, b_7 is estimated by b_6U, b_8 by b_6U^2 and so on. Approach A effectively involves linear prediction on a log scale. Approach B investigated whether predictions could be improved by using more values than the last two and allowing for non-linearity. We used a quadratic model fitted to the log of all six period values, weighting these values by powers of 2, from 1 for period 1 to 32 for period 6.

The APC model fitted to the 1961-1990 data was used to produce predictions for 1991-2010, which could be validated against the known data for that period. The model fitted to the 1981-2010 data was used to produce predictions for 2011-2030, outside the range of the known data.

2.4 Validating the predictions and comparing the approaches

Having used 1961-1990 data to generate predictions for 1991-2010, these predictions were then compared with known data, and approximate chi-squared goodness-of-fit statistics estimated as described in section 2.2. Goodness-of-fit statistics based on predictions of the APC model were also compared with those based on the age model (A; see section 2.2) to evaluate how much allowance for period and cohort effects improved prediction. Also, for each five year time period, observed and predicted deaths were each summed over the age groups, with the ratio of the two totals (Q = total observed/total predicted) used as an indicator of misfit, a value of Q of 1 implying that the period value was correctly estimated. The product of Q and the estimated period value estimates what the period value should have been had the prediction been accurate.

For each disease and sex, and for UK and Switzerland only, the relative merits of Approaches 1, 2 and 3 for extrapolating cohort values were assessed by comparing goodness-of-fit statistics based on the four extrapolation time periods.

For each disease and sex, and for each country (except Germany) the relative merits of Approaches A and B (see section 2.3) for extrapolating period values were also assessed by comparison of goodness-of-fit statistics. Comparisons were also made based on the number of estimates using each approach where Q was within ±5% of 1, within ±10%, ±25%, ±50% or ±100%, and outside ±100%, giving scores of 5, 4, 3, 2, 1 and 0 for each successive category.

2.5 Relating cohort effects to smoking habits

Formal modelling relating disease rates to smoking habits was not attempted. However cohort values derived from the 1981-2010 data were correlated with cohort-based estimates of cigarette consumption per person at ages 30-34, 40-44 and 50-54, where available from the International Mortality and Smoking Statistics database [16].

2.6 Software

Analyses were conducted using in-house software written in Intel Visual Fortran V9, developed from a version used in a previous publication [17]. Output for lung cancer in UK females was checked against results given by Osmond and Gardner [2].

3 Results

3.1 Fitting the APC model to data for 1961-1990

Table 2 summarizes aspects of fit of the APC model to the 1961-1990 data. Various broad conclusions can be drawn. Firstly, the first of the four blocks in Table 2 shows that the percentage of variance from the A model explained by the APC model is nearly always high, exceeding 99% in 35 of 88 analyses, and exceeding 90% in all but three. The smallest percentage explained for COPD is 72.3% for females in Switzerland, for lung cancer is 87.9% for females in Austria, for IHD is 89.1% for females in Hungary, and for stroke is 96.1

for males in Austria. Despite the APC model typically explaining a large percentage of the variance in nearly all cases, it is evident from the last block that the fit is usually not perfect, the goodness-of-fit chi-squared on 32 d.f. exceeding 67.2 ($p<0.001$) in nearly all analyses for IHD and stroke, and about half those for lung cancer and COPD, and being over 1000 in two analyses. The second and third blocks show the percentages of variance explained respectively by cohort and by period. For lung cancer in both sexes, and stroke in females, cohort explains a higher percentage than period in 31 of 33 analyses. However, for COPD in females and stroke in males, period explains a higher percentage in 16 of 22 analyses. In the remaining combinations, no consistent pattern is evident, cohort explaining more than period in 15 of 33 analyses.

Figures 2 and 3 show examples of the fitted period and cohort values. In Figure 2 (lung cancer, Austria, males) cohort on its own explained a large part (96.2%) of the variance, whereas period explained only a small part (19.3%). In Figure 3 (IHD, Switzerland, females) the reverse was true, cohort explaining 34.6% and period 93.5%.

Figures 4 and 5 illustrate the goodness-of-fit for the two cases with the smallest percentage of variance explained by the APC model (COPD, Switzerland, females; lung cancer, Austria, females), data being shown only for alternate age groups. Given these are the worst cases, the fit to the APC model would appear overall to be relatively good.

For each analysis, Supplementary File 1 gives fuller details, not only of the fit of the APC model to 1961-1990 data, but also of the predictions for 1991-2000 (using Approach 3 for period values and Approach A for cohort values).

3.2 Comparing observed data for 1991-2010 with values predicted by extrapolation

A first series of analyses used Approach A for extrapolating period values, comparing results using Approaches 1, 2 and 3 for extrapolating cohort values, based on data for UK and Switzerland. Table 3 shows the goodness-of-fit statistics using the three approaches. Compared to Approach 1, the other approaches fitted better (lower value) in most cases; 14 of 16 for Approach 2 and 12 of 16 for Approach 3. Comparing Approaches 2 and 3, no consistent difference was seen. Thus, ignoring cases where the goodness-of-fit statistics for the two approaches were within 5%, Approach 3 was somewhat better in five cases (COPD in

UK females, and lung cancer and COPD in Switzerland in both sexes) and somewhat worse in five cases (lung cancer in the UK in both sexes, COPD in UK males, and stroke in Switzerland in both sexes). Because other work had been carried out using Approach 3, and as effects of cohort extrapolation are minor (only affecting younger age groups with relatively few deaths) compared to effects of period extrapolation, it was decided to use Approach 3 for further work.

Using Approach 3 for cohort values, comparison was then made of results using Approaches A and B for extrapolating period values (Table 4). For lung cancer in males, Approach B (quadratic extrapolation) was better than Approach A (ratio of last two period values), producing a lower goodness-of-fit chi-squared in 8 of 11 countries and having a higher (better) total score (33 vs 24) based on closeness to 1 of the observed/predicted ratio for 2006-2010. However, this was an exceptional case, with the total score over all data sets being 202 using Approach A and 173 using Approach B, with the chi-squared value lower with Approach A in 51 of 88 (58.0%) comparisons. Notably, there were 36 cases where the goodness-of-fit chi-squared was over twice as large for Approach B as for Approach A, as against only 19 where the reverse was true. The difference between the approaches was particularly marked for COPD, where the fit was often substantially worse using Approach B.

Figures 6 to 7 illustrate predictions using the two approaches, showing the fitted period values for 1961-1990 and the predicted values for 1991-2010 under each approach, as well as the "true" period values; those which would have been necessary for the observed/predicted to equal 1. The full set of 88 figures is available in Supplementary File 1.

Figure 6, for COPD in females for USA, illustrates the majority of cases where the estimates seem quite accurate. Here observed/predicted values for 2006-2010 are 0.915 under Approach A and 0.948 under Approach B. However, some predictions are very inaccurate. In Figure 7, for COPD in females for Sweden, the observed/predicted values for the final period are 1.864 for the two-period linear model and 8.818 for the quadratic model, it being clear visually that the true values for 1991-2010 could not reasonably have been predicted using the 1961-1990 data.

Because Approach B had no overall advantage on any of the tests made, and was substantially worse than Approach A regarding gross misfits, it was decided to use Approach A for extrapolating period effects (and Approach 3 for cohort effects) for further work.

Table 5 summarizes aspects of fit of the predictions to 1991-2010 data based on these approaches. The goodness-of-fit chi-squared statistics are, unsurprisingly, substantially greater for the prediction period than those in Table 2 for the period (1961-1990) to which the model was fitted. Nevertheless, in the great majority of cases, prediction using the APC model produced a substantially lower chi-squared than did prediction using the A model, the "variance explained" by period and cohort exceeding 90% in 46 of the 88 analyses (52.3%), and 80% in a further 12 (13.6%). However, the APC model does not always improve predictions, and in 11 cases (12.5%) actually predicts less well than the A model, illustrated by the negative values in Table 5.

Fuller details of these projections are also in Supplementary File 1.

3.3 Fitting the APC model to data for 1981-2010

Table 6 summarizes aspects of the fit of the APC model to the 1981-2010 data. As for the 1961-1990 data (Table 2), the percentage of residual variance explained by the APC model is nearly always high, exceeding 99% in 53 of 96 analyses, and 90% in all but five. The smallest percentage explained, 70.9%, is again for COPD. Although the APC model explains a large proportion of the variance, there are again many occasions where the goodness-of-fit chisquared exceeds 67.2 ($p < 0.001$), though the proportion (56.3%) is less than for Table 2 (71.5%). For both the 1961-1990 and 1981-2010 data, the percentage of variance explained by cohort was usually higher than that explained by period, but whereas this was true for 59.1% of the analyses in Table 2, it was true for 82.3% of those in Table 6, 11 of the 17 exceptions being for stroke.

Supplementary File 2 provides additional information, laid out similarly to that for 1961-1990.

3.4 Predictions for 2011-2030

Table 7 presents observed rates for 1981-1985 and 1996-2000, and rates predicted by the APC model for 2011-2016 and 2026-2030 for the combined age group 30-79, standardized to the European 1976 population [18].

For lung cancer in males, rates for North America and Western Europe show a marked decline over the whole period, predicted 2026-2030 rates being less than 40% of those in 1981-1985, except for France. Rates in Eastern Europe, though not smoothly declining, are still predicted to be markedly lower in 2026-2030 than 1981-1985. A smaller decline is predicted for Japan. For many countries, rates in females are predicted to rise, to be not dissimilar to those in males by 2026-2030. For Canada, Japan, Sweden and USA, however, the rates are predicted to be lower in 2026-2030 than they were during the observed period.

For COPD, rates in males also often show a striking decline, with rates in 2026-2030 less than 35% of those in 1981-1985 in 10 countries. The decline is much less than this for the USA, with an increase predicted for Hungary. In females, rates are also generally predicted to decline, though not in USA or Hungary.

For IHD in both sexes, 10 of 12 countries show a striking decline, with rates in 2026-2030 typically only 10 to 20% of those in 1981-1985. A decline is also predicted for Hungary and Japan, though less markedly.

The pattern for stroke is similar to that for IHD. Here the smallest decline is for males in Poland, where the 2026-2030 rate is predicted to be 44% of that in 1981-1985. In all other countries they are predicted to be less than 25%, the decline often being more than this.

Fuller details of the projections are in Supplementary File 2.

3.5 Relationship of fitted cohort values to cigarette consumption

Table 8 presents correlation coefficients based on relating cohort values fitted to the 1981-2010 data (using the APC model) with estimates of cigarette consumption per person at age 30-34 for the same cohort. The correlations are generally higher for males, and highest for

lung cancer. For lung cancer, correlations exceed 0.50 in 21 of 24 cases in both sexes, and 0.80 in 13. For IHD and stroke, however, correlations in females were as often negative as they were positive. Results based on cigarette consumption at ages 40-44 and 50-54 were quite similar, the mean correlations being given in the Table 8 footnotes.

Figures 8 to 11 show examples, one per disease, of strong cohort effects in conjunction with a relatively high correlation with average cigarette consumption.

4 Discussion

As a method for summarizing age-specific mortality data over a period, the APC model has been quite widely used (e.g. [3, 13]). The analyses presented here confirm its usefulness for the major smoking-related diseases, whether based on data for 1961-1990 or 1981-2010. For the great majority of the countries and diseases studied, the APC model explained a very large proportion of the residual variance. Though a few cases explain less than 90%, even the worst (see Figures 4 and 5) give a visually not unreasonable fit, taking account of the relatively few cases at age 30-34. The analyses also show the importance of taking cohort into account. This is emphasised partly as the percentage of variance explained by cohort usually exceeds that explained by period (particularly for 1981-2010), and partly as, for the most strongly smoking-related disease, lung cancer, there is a consistent correlation between fitted cohort values and cohort-specific estimates of cigarette consumption. For IHD and stroke, for which smoking is only one of various major risk factors, and where considerable treatment advances have occurred, it is unsurprising that the correlations are poorer, particularly for females.

As expected, given that statistics students are often advised not to extrapolate beyond the range of their observations and that government economists sometimes fail to predict accurately even a few months ahead, the APC model is clearly poorer at predicting future rates than describing current rates. To carry out such predictions involves extrapolating both cohort and period values. We compared three approaches for cohort extrapolation, each involving log-linear extrapolation using weights decreasing exponentially in the past, and found that omitting either the final value or the final two values from the extrapolation were similarly accurate for predicting known data for 1991-2010, and better than including all the values, including the final one which is based on data only for one period for age 30-34, often

with few deaths. We decided to use approach 3, omitting only the last value, estimating it and the future cohort values by extrapolation. The choice of approach only affects results for younger age groups where relatively few deaths occur, so is not vital to the predictions, and is one reason why we only compared approaches for two countries.

The effect of extrapolating period values is much more important. Approach A (ratio of last two period values) was selected in preference to Approach B (quadratic extrapolation), as Approach A was less likely to produce gross misfits. Often (e.g. Figure 6), the predictions for 1991-2010 from 1961-1990 rates were quite close to the actual data, but there were some obvious exceptions. In some cases (e.g. Figure 7), it is difficult to envisage any model not including data on trends in risk factors, treatment or diagnosis which could adequately explain observed trends in rates.

Though there are clearly limitations to any model which predicts future rates based only on recent mortality trends, we consider the APC model predictions shown in Table 7 (with further detail given in Supplementary File 2) quite plausible. Certainly, there seems no obvious reason why the sharply declining trends in IHD and stroke should not continue, so that by 2026-2030 rates will be very much lower than in 1981-1985, making these diseases largely causes of deaths in the 20th Century.

5 Acknowledgements

We thank WHO and UN for making the mortality and population data available. We also thank Barbara Forey for helpful comments on a draft, and for her contributions to the work conducted in the 1980s which we relied upon. Chris Gooch also conducted much of the earlier work, and we thank him, as well as Pauline Wassell, Diana Morris and Yvonne Cooper, for assistance in typing various drafts of the paper and obtaining relevant literature.

6 Funding

This work was supported by Philip Morris Products S.A. whom we also thank.

7 Supplementary Files

1) APC_Supplementary_File_1.pdf
2) APC_Supplementary_File_2.pdf

The supplementary files are available for download from the following links:-

http://www.pnlee.co.uk/documents/refs/APC_Supplementary_File_1.pdf

http://www.pnlee.co.uk/documents/refs/APC_Supplementary_File_2.pdf

8 Figures

Figure 1. Ilustrative diagram of age group, periods and cohorts used for predicting data for 2011-2030 from data for 1981-2010

The numbers in the body of the diagram identify the cohorts. The cohort numbers (k) are related to the age groups (i) and the periods (j) by the formula k = j – i + 10. The cohort numbers in red represent cohorts where we have no mortality data. The modelled values for earlier cohorts were used to provide predicted values for these cohorts.

		Period							Predicted period		
		j =1	2	3	4	5	6	7	8	9	10
Age group		1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010	2011-2015	2016-2020	2021-2025	2026-2030
i = 10	75-79	1	2	3	4	5	6	7	8	9	10
9	70-74	2	3	4	5	6	7	8	9	10	11
8	65-69	3	4	5	6	7	8	9	10	11	12
7	60-64	4	5	6	7	8	9	10	11	12	13
6	55-59	5	6	7	8	9	10	11	12	13	14
5	50-54	6	7	8	9	10	11	12	13	14	15
4	45-49	7	8	9	10	11	12	13	14	15	16
3	40-44	8	9	10	11	12	13	14	15	16	17
2	35-39	9	10	11	12	13	14	15	16	17	18
1	30-34	10	11	12	13	14	15	16	17	18	19

Figure 2. Period (of death) and (birth) cohort values fitted to 1961-1990 data for lung cancer, Austria, males

Cohort explains a large part of the variance while period explains only a small part. Note that cohorts and periods are labelled by the midpoint year.

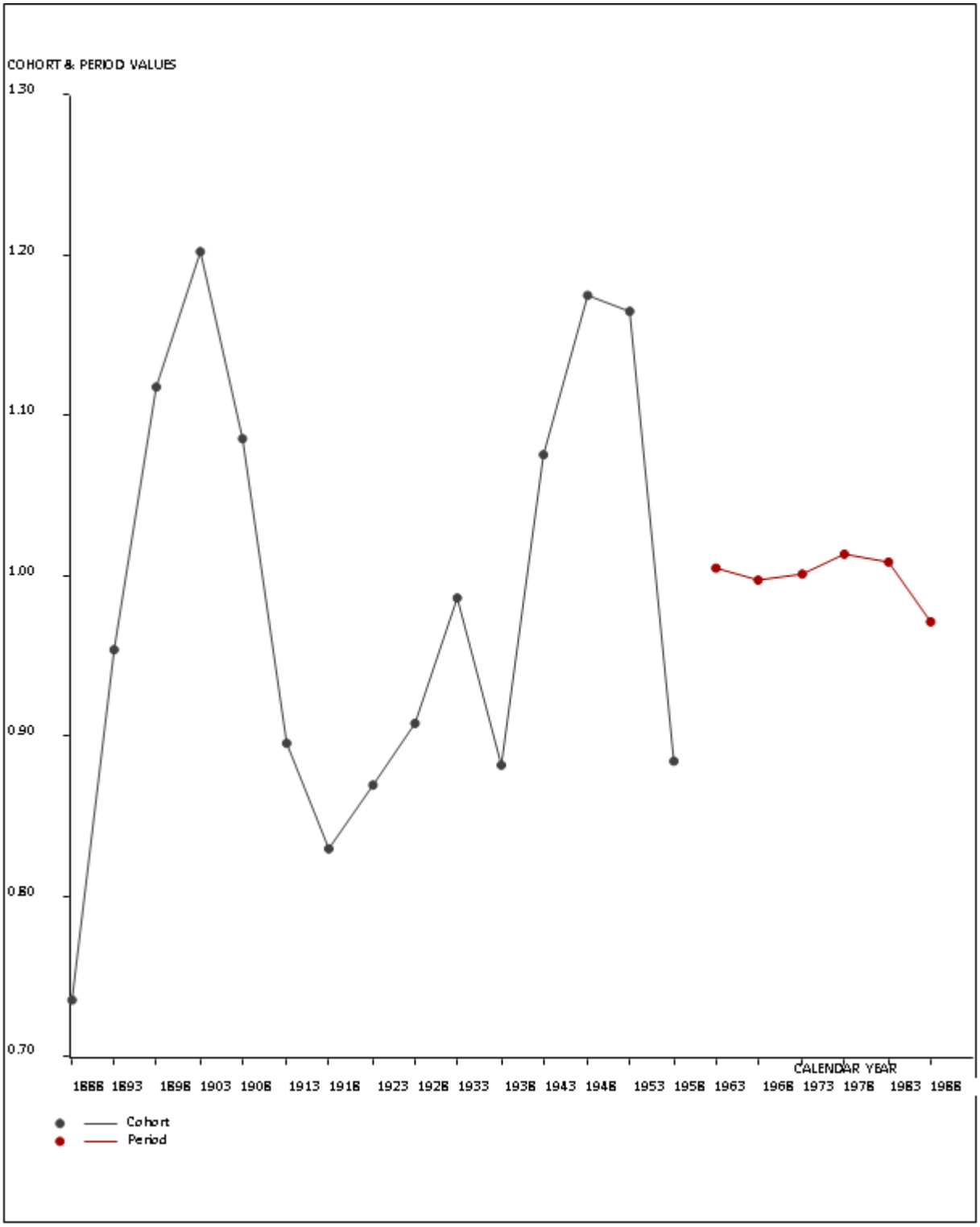


Figure 3. Period (of death) and (birth) cohort values fitted to 1961-1990 data for IHD, Switzerland, females

Period explains a large part of the variance while cohort explains only a small part.

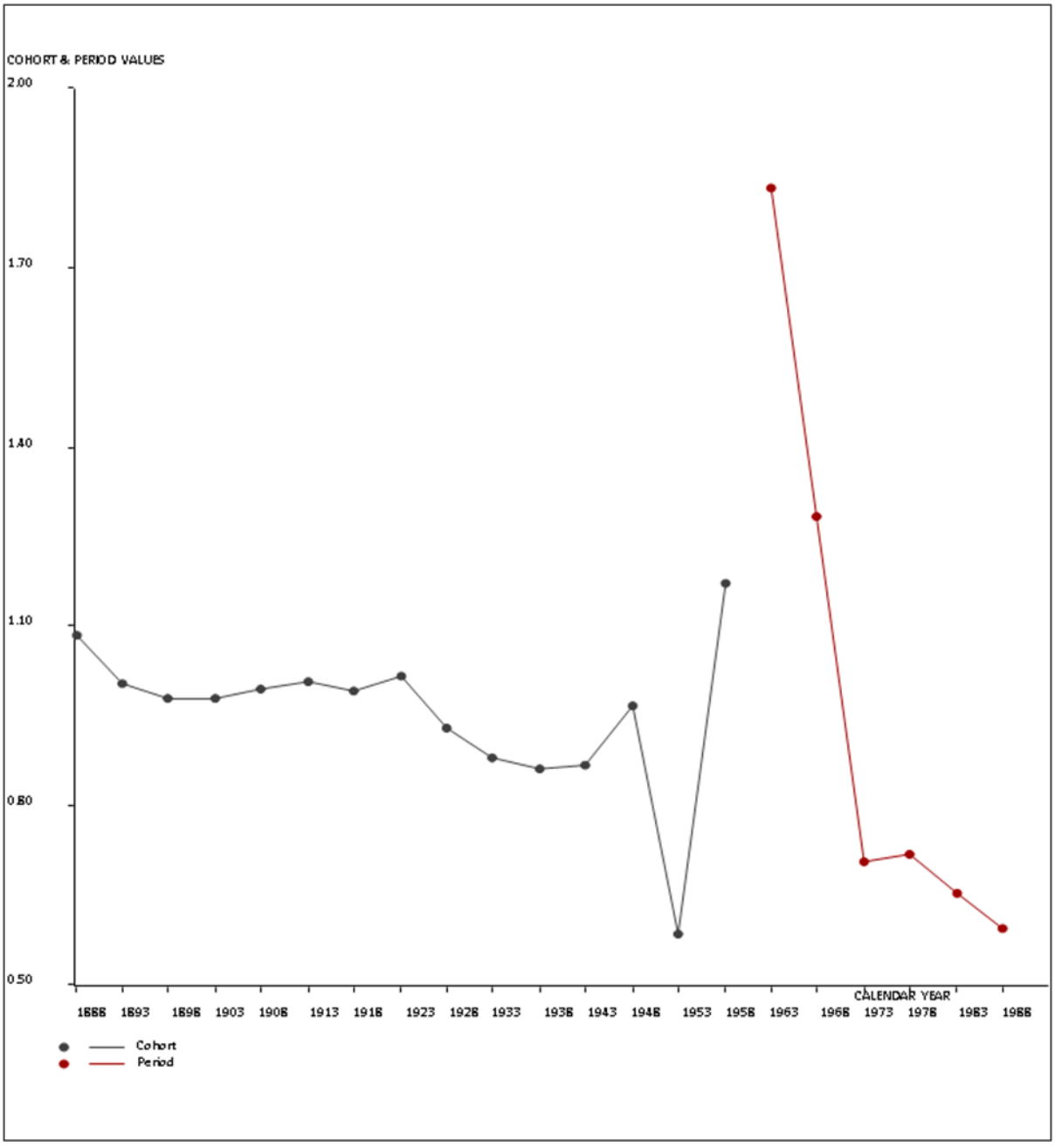


Figure 4. Observed and fitted mortality rates (1961-1990) for selected age groups; COPD, Switzerland, females

Only 72.3% of variance is explained by the APC model but the fit appears to be reasonable.

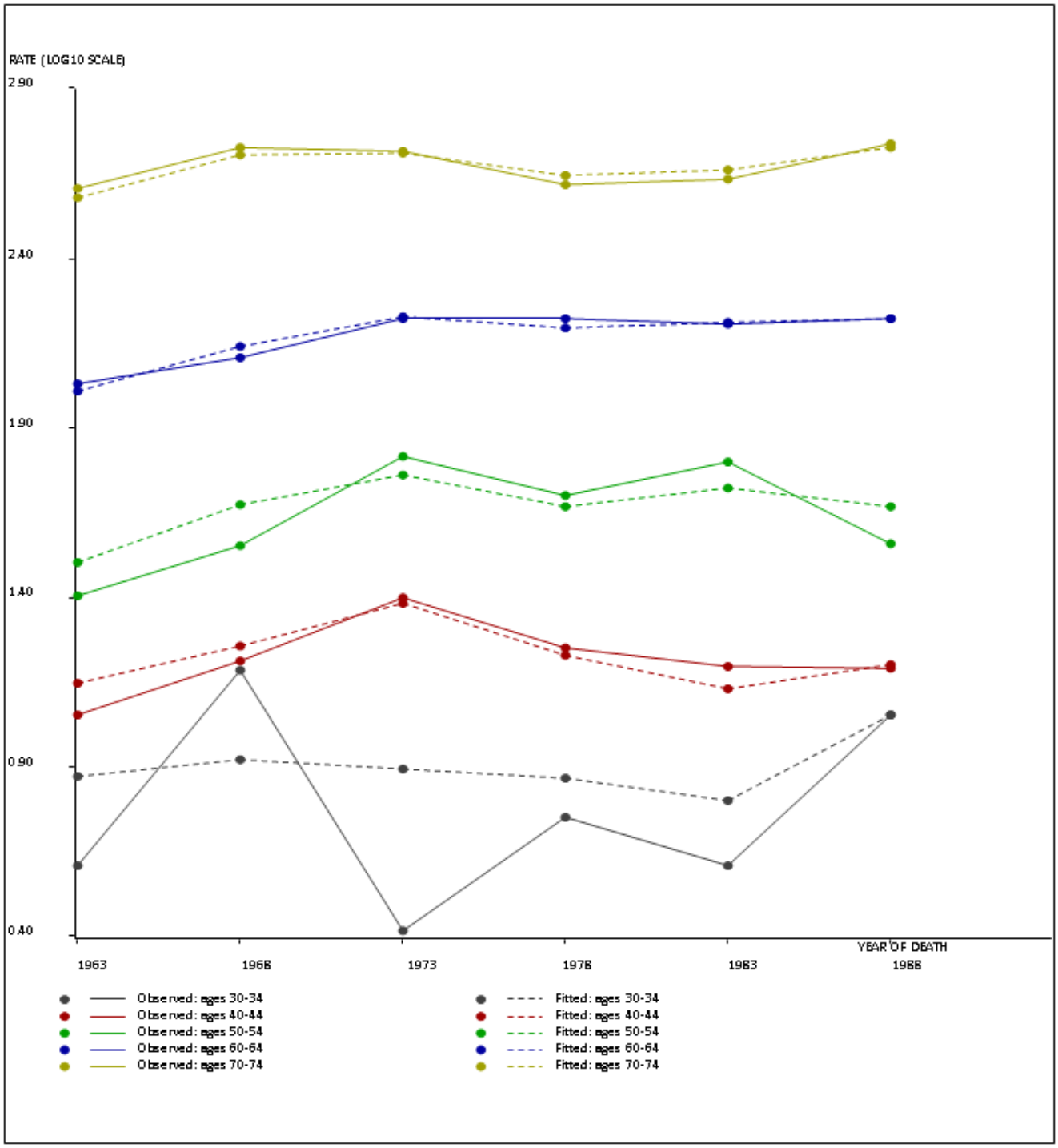


Figure 5. Observed and fitted mortality rates (1961-1990) for selected age groups; lung cancer, Austria, females

Only 87.9% of variance is explained by the APC model but the fit appears to be reasonable.

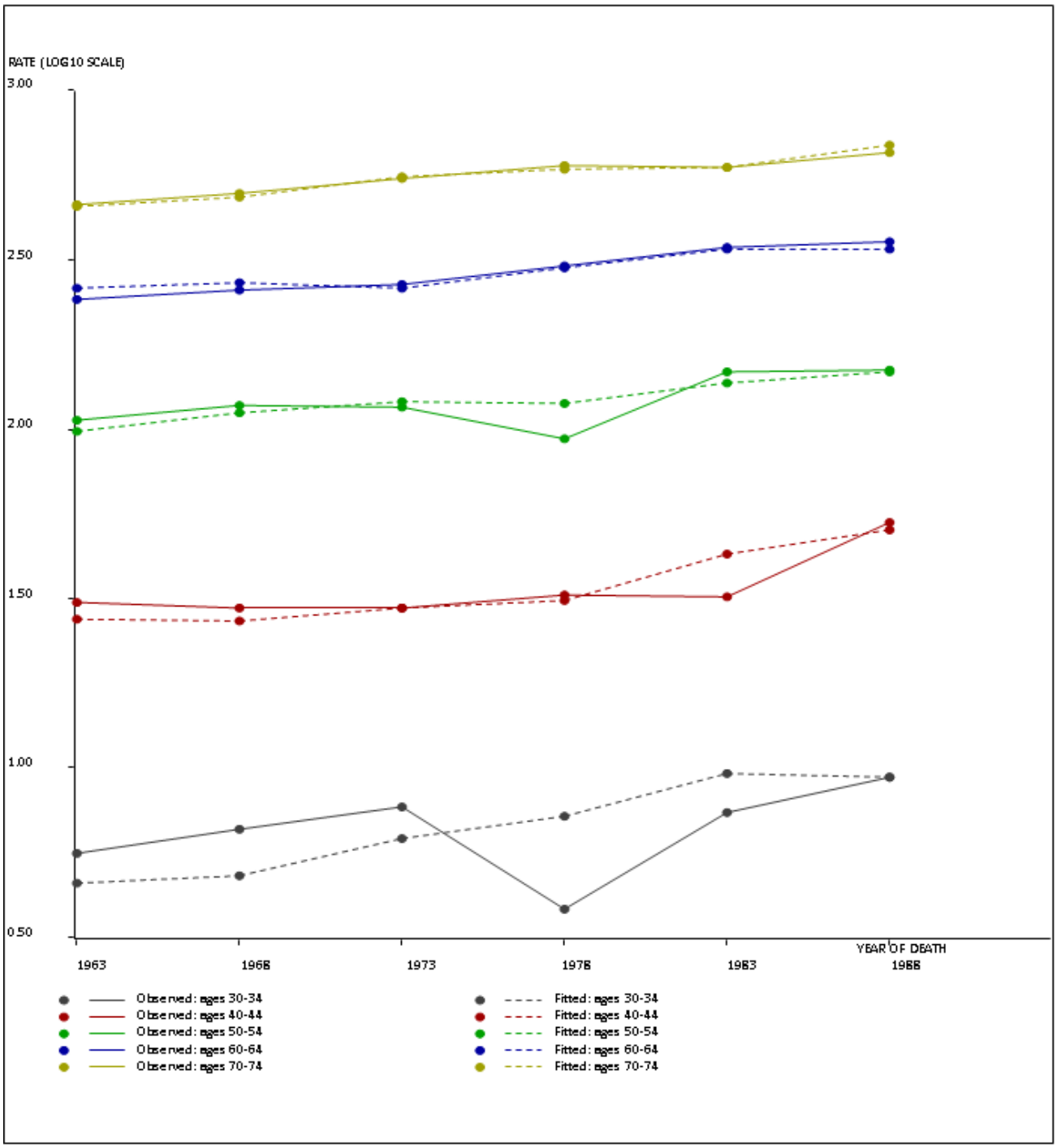


Figure 6. Comparing the two approaches used to predict period effects for 1991-2010 based on data for 1961-1990: COPD, USA, females

This example shows Approaches A and B giving similar results, both close to the true values.

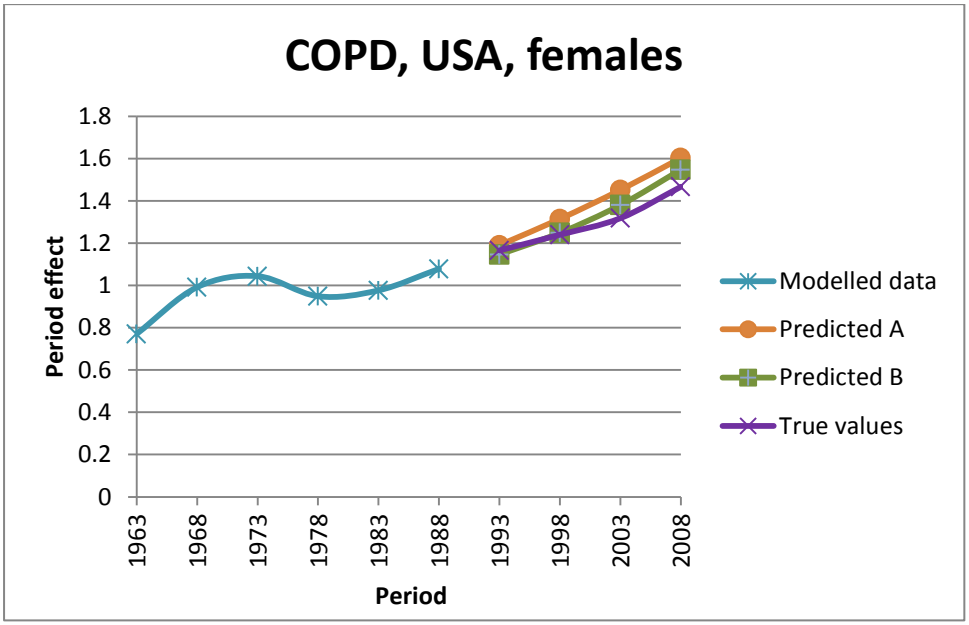


Figure 7. Comparing the two approaches used to predict period effects for 1991-2010 based on data for 1961-1990: COPD, Sweden, females

This example shows neither of Approaches A or B predicting the true values accurately.

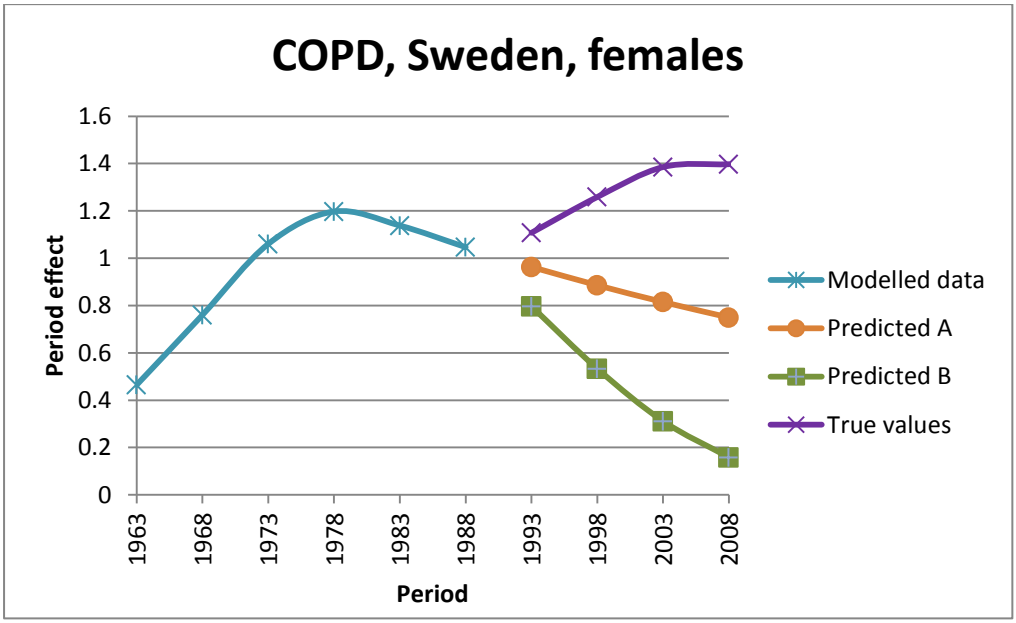


Figure 8. Cigarette consumption (average cigarettes per person) and fitted cohort values; IHD, Canada, males

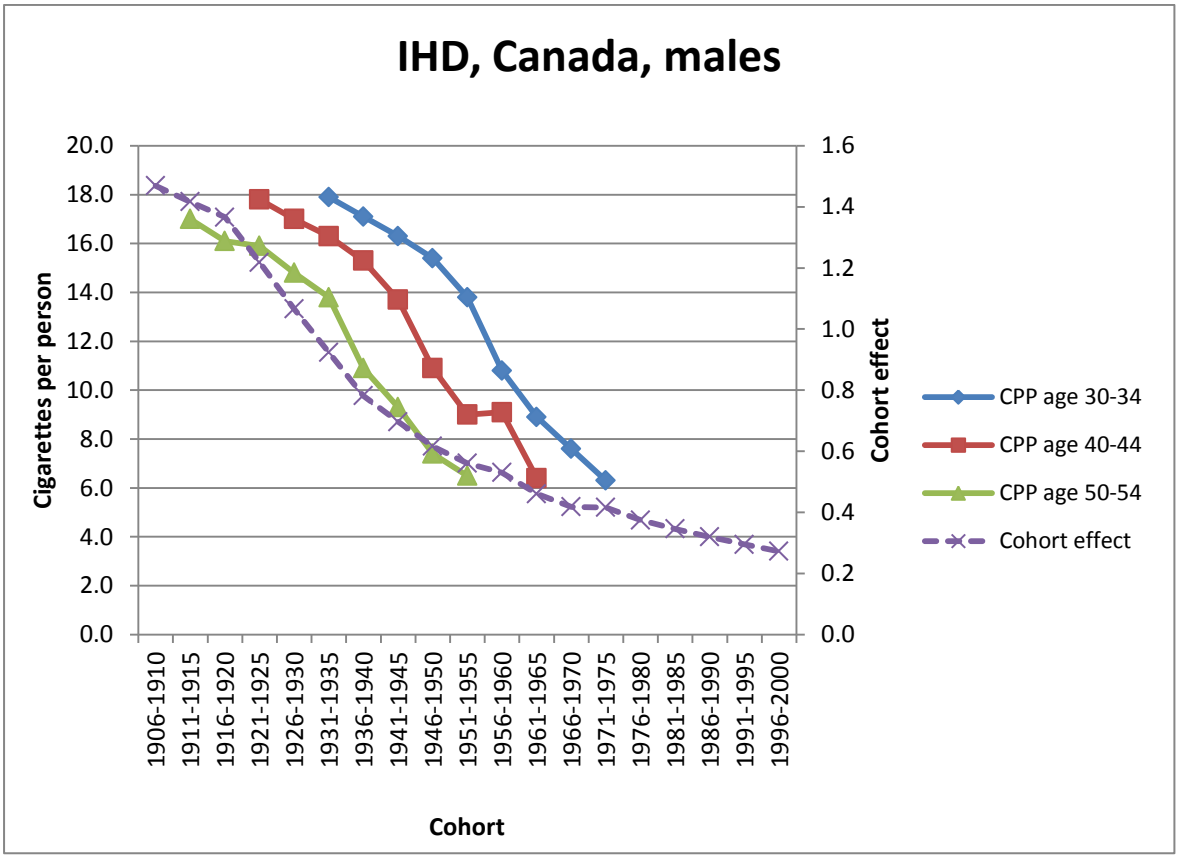


Figure 9. Cigarette consumption (average cigarettes per person) and fitted cohort values; stroke, France, males

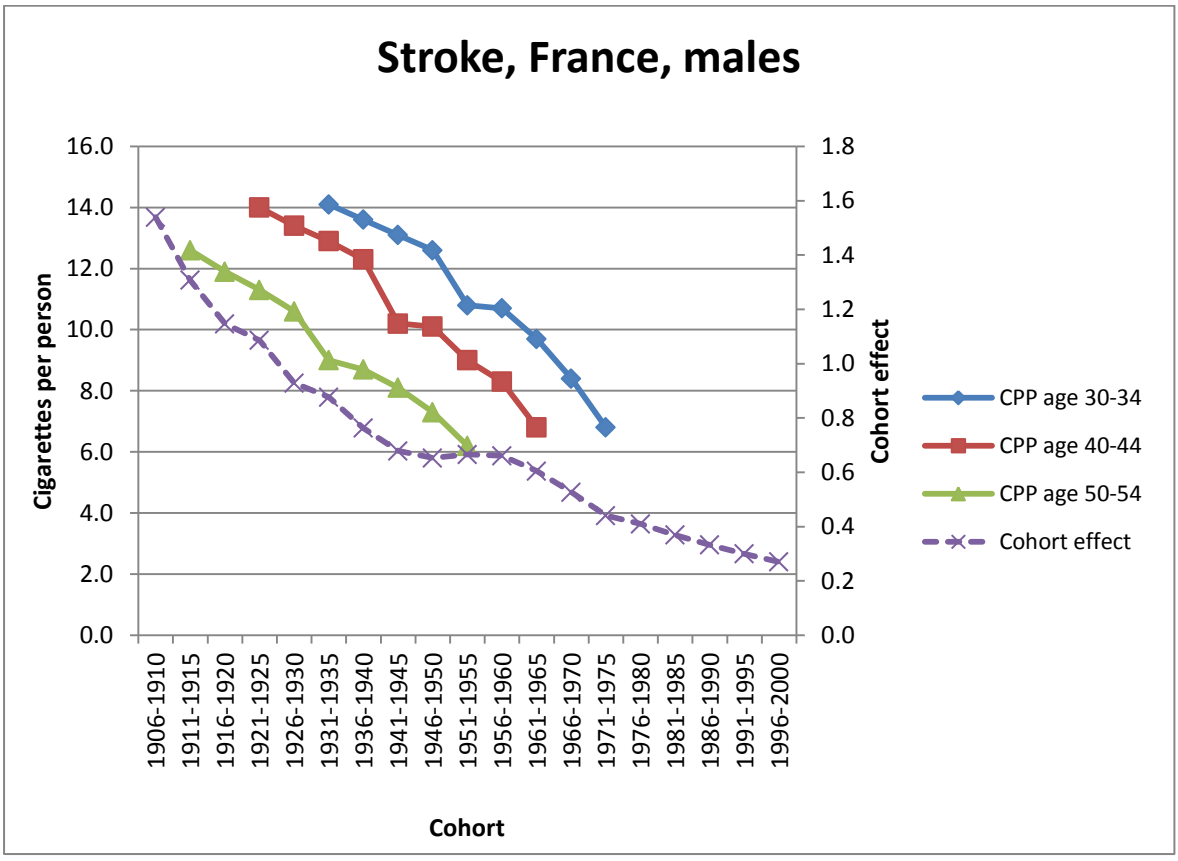


Figure 10. Cigarette consumption (average cigarettes per person) and fitted cohort values; lung cancer, France, females

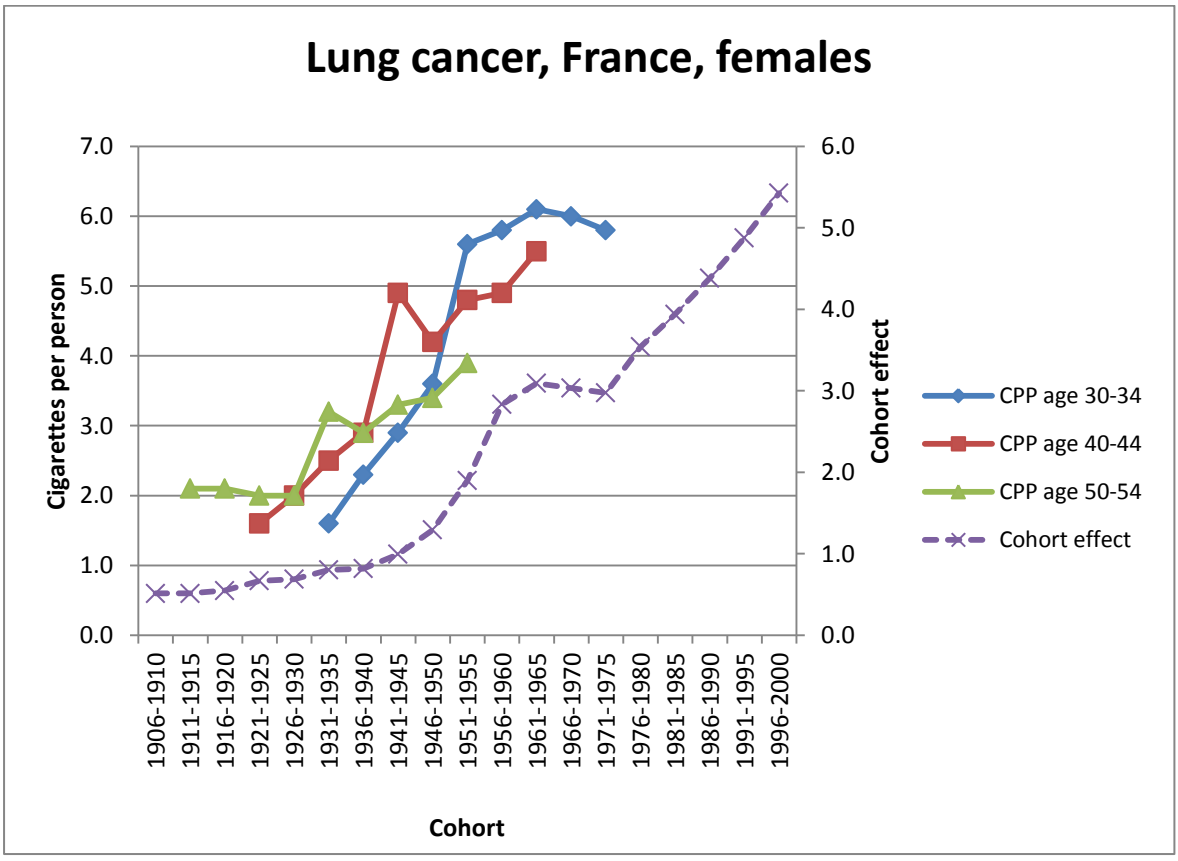
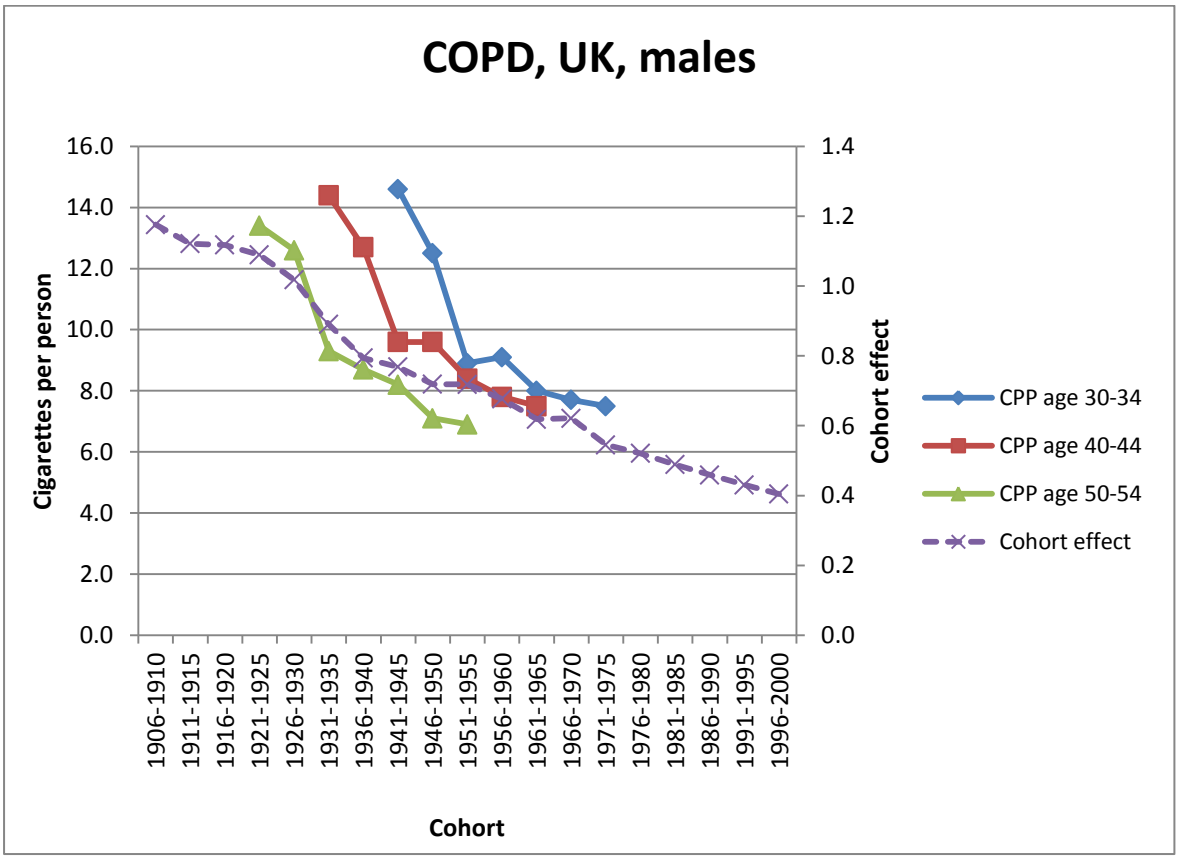


Figure 11. Cigarette consumption (average cigarettes per person) and fitted cohort values; COPD, UK, males



9 Tables

Table 1. Years in which the countries used each ICD revision

	ICD Revision				
Country	7	8	9	10	Switzerland[1]
Austria	1955-1968	1969-1979	1980-2001	2002-2010	-
Canada	1951-1968	1969-1978	1979-1999	2000-2009	-
France	1952-1967	1968-1978	1979-1999	2000-2009	-
Germany[2]	-	-	1980-1997	1998-2010	-
Hungary	1955-1968	1969-1978	1979-1995	1996-2010	-
Italy[3]	1951-1967	1968-1978	1979-2002	2003-2010	-
Japan	1951-1967	1968-1978	1979-1994	1995-2010	-
Poland[4]	1959-1968	1969-1979	1980-1996	1999-2010	-
Sweden	1951-1968	1969-1986	1987-1996	1997-2010	-
Switzerland	1951-1968	1969-1994	-	-	1995-2010
UK[5]	1951-1967	1968-1978	1979-2000	2000-2010	-
USA	1951-1967	1968-1978	1979-1998	1999-2010	-

[1] Switzerland used its own codings instead of ICD revisions 9 and 10 (see text).
[2] No data were available for 1951-1979.
[3] No data were available for 2004 and 2005.
[4] No data were available for 1997 and 1998.
[5] In 2000 Scotland used ICD revision 10 while the other parts of the UK used ICD revision 9.

Table 2. Fitting the APC model to data for 1961-1990

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
Percentage of variance explained by period and cohort[1]								
Austria	97.1	87.9	98.0	93.0	93.5	97.5	96.1	98.1
Canada	99.2	99.4	95.0	96.0	99.5	99.5	99.4	99.2
France	99.1	93.0	99.2	99.2	91.4	98.0	99.7	99.7
Hungary	99.4	96.3	98.6	95.0	95.7	89.1	98.5	95.1
Italy	99.2	98.6	98.9	99.1	99.4	99.9	97.6	98.4
Japan	99.5	98.5	98.4	97.4	98.8	98.5	99.8	100.0
Poland	99.7	96.2	97.5	96.3	97.9	95.9	98.7	96.5
Sweden	96.9	97.7	95.0	95.3	97.6	99.6	96.3	97.8
Switzerland	94.8	94.7	92.3	72.3	97.3	99.4	99.3	99.2
UK	99.8	99.8	99.2	96.1	92.7	96.6	98.7	99.0
USA	99.5	99.9	98.4	99.4	99.9	99.9	99.9	99.9
Percentage of variance explained by cohort[2]								
Austria	96.2	48.0	82.0	65.4	72.6	48.5	55.2	74.5
Canada	94.5	86.0	92.0	90.6	87.0	46.2	47.1	51.5
France	89.5	54.7	90.2	87.8	71.2	91.5	91.5	91.4
Hungary	97.2	71.8	88.2	68.6	94.0	63.0	97.2	91.6
Italy	95.8	88.0	97.2	51.1	93.4	80.0	41.3	67.2
Japan	96.6	95.0	79.7	30.9	70.8	87.7	95.3	98.8
Poland	92.6	49.0	78.9	86.8	95.8	88.7	90.8	78.4
Sweden	90.5	75.6	47.1	77.5	54.0	82.1	27.1	30.4
Switzerland	88.9	31.2	85.4	51.1	67.8	34.6	62.8	77.7
UK	99.7	98.8	94.5	93.5	73.8	85.4	38.2	43.7
USA	98.1	99.3	97.5	97.1	93.2	85.0	93.4	88.5
Percentage of variance explained by period[3]								
Austria	19.3	10.1	96.4	87.3	71.6	65.9	70.2	52.7
Canada	75.4	36.0	75.9	71.3	85.4	65.1	84.4	43.7
France	69.6	74.3	96.8	96.2	75.7	60.1	96.4	87.4
Hungary	58.8	35.1	94.3	86.9	68.9	27.2	94.4	88.5
Italy	67.8	50.6	83.3	80.8	94.1	95.3	56.7	32.0
Japan	44.1	43.9	96.5	94.3	50.0	68.7	92.7	96.6
Poland	90.2	11.0	96.0	94.1	65.1	82.6	95.6	94.9
Sweden	81.8	12.7	88.4	87.6	96.8	94.7	28.8	29.6
Switzerland	50.5	12.9	75.2	59.9	75.8	93.5	59.5	34.6
UK	68.1	25.6	80.8	76.7	77.2	63.4	47.9	20.6
USA	82.0	89.1	90.4	77.7	98.4	97.8	97.8	95.5

Table 2 continued

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
Goodness-of-fit chisquared for the APC model (on 32 d.f.)[4]								
Austria	35.6	37.9	40.9	50.3	115.8	102.0	110.3	100.5
Canada	47.0	51.4	161.6	80.4	198.9	112.4	71.7	128.1
France	121.5	59.7	123.7	65.9	400.3	117.5	117.8	172.1
Hungary	49.2	52.4	50.0	77.0	290.3	521.2	70.5	103.8
Italy	242.5	39.1	225.8	102.1	232.0	104.2	747.1	653.1
Japan	105.4	68.6	205.3	195.2	612.6	1159.9	750.0	140.4
Poland	72.4	111.7	115.9	83.4	425.9	421.4	74.5	98.5
Sweden	30.4	27.2	58.5	34.2	72.6	22.9	97.9	135.6
Switzerland	51.2	31.4	44.0	45.0	154.6	93.3	35.2	58.7
UK	34.5	48.9	380.0	179.1	1118.8	430.3	391.4	410.6
USA	237.4	82.7	326.0	339.80	547.8	396.4	144.2	178.0

[1] 100 (RSS_A - RSS_{APC}) / RSS_A where RSS_A is the residual sum of squares for model A and RSS_{APC} is that for model APC. Critical values are 53.3, 60.1 and 67.2 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[2] 100 (RSS_{AP} - RSS_{APC}) / RSS_{AP} where RSS_{AP} is the residual sum of squares for model AP. Critical values are 45.4, 52.8 and 60.8 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[3] 100 (RSS_{AC} - RSS_{APC}) / RSS_{AC} where RSS_{AC} is the residual sum of squares for model AC. Critical values are 25.1, 33.2 and 43.1 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[4] Critical values are 46.2, 53.5 and 62.5 for, respectively $p<0.05$, $p<0.01$ and $p<0.001$.

Table 3. Goodness-of-fit statistics[1] to the APC model (for prediction from 1961-1990 to 1991-2010) using three different approaches[2] for extrapolating cohort effects

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
UK								
Approach 1	3030	2924	5824	3028	67648	71611	10282	7046
2	936.1	1443	4135	2675	66948	71104	10109	6749
3	1348	2222	5024	2539	67362	71648	10106	6951
Switzerland								
Approach 1	407.8	214.4	1307	838.5	2215	1395	396.1	728.8
2	534.2	285.1	1211	641.0	1942	1347	266.7	699.7
3	486.9	237.7	1145	610.0	1961	1376	309.6	802.7

[1] Approximate chisquared values based on summing (observed-expected)^2 / expected over each of the cells in the prediction period.
[2] See methods section 2.3 for definition of the approaches studied.

Table 4. Comparison of Approach A and Approach B for estimating period effects[1] (using 1961-1990 data to predict data for 1991-2010)

	Lung cancer		COPD		IHD		Stroke		
	Males	Females	Males	Females	Males	Females	Males	Females	Total
Chisquared fit better[2]									
Approach A	3	8	10	8	5	6	5	6	51
Approach B	8	3	1	3	6	5	6	5	37
Chisquared fit much better[3]									
Approach A	2	3	7	6	5	4	5	4	36
Approach B	5	1	1	2	2	3	4	1	19
Error in prediction[4] for final period for Approach A									
<5%	0	1	0	0	0	0	3	1	5
5% to <10%	1	3	0	1	2	1	1	0	9
10% to <25%	3	4	7	4	1	3	1	3	26
25% to <50%	5	2	1	2	5	1	2	4	22
50% to <100%	1	1	3	3	1	3	4	3	19
100% or more	1	0	0	1	2	3	0	0	7
Total score	24	34	26	23	22	18	30	25	202
Error in prediction[4] for final period for Approach B									
<5%	2	0	0	1	0	1	0	0	4
5% to <10%	2	2	0	1	0	0	3	0	8
10% to <25%	3	2	2	1	4	3	2	3	20
25% to <50%	2	5	3	1	0	2	3	5	21
50% to <100%	2	1	1	3	5	3	2	2	19
100% or more	0	1	5	4	2	2	1	1	16
Total score	33	25	13	17	17	21	26	21	173

[1] In Approach A, period values were extrapolated based on the ratio of the last two period values. In Approach B, a quadratic model was fitted to the log of all six period values, weighted by powers of 2 from 1 for period 1 to 32 for period 6.
[2] Based on a comparison of goodness-of-fit chisquared value for the APC model for the prediction period.
[3] An approach was considered much better if its goodness-of-fit chisquared was less than half that of the other approach.
[4] Based on the closeness to 1 of Q, the ratio total observed/total predicted. Scores of 5, 4, 3, 2, 1 and 0 are given for the six categories.

Table 5. Comparing observed data for 1991-2010 with that predicted by the APC model fitted to data for 1961-1990 using Approach A for extrapolating period values and Approach 3 for extrapolating cohort values

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
Percentage of variance explained by period and cohort[1]								
Austria	48.6	91.2	82.1	-124.4	94.9	86.9	95.0	96.7
Canada	49.5	88.4	96.6	81.7	99.4	99.1	97.4	97.4
France	-196.0	81.3	91.7	50.7	93.3	91.2	99.2	98.4
Hungary	-60.4	86.5	41.0	5.2	-245.5	48.1	20.2	60.3
Italy	58.1	88.1	98.8	97.8	98.8	99.5	97.5	96.6
Japan	16.8	31.0	96.8	94.0	-83.0	15.2	90.5	97.1
Poland	31.4	91.6	88.4	92.7	-847.5	-688.6	-65.8	-187.5
Sweden	49.8	96.5	-81.6	14.0	96.1	96.8	94.6	96.4
Switzerland	89.3	97.1	76.5	-108.2	94.4	94.8	98.7	96.8
UK	99.0	94.5	97.0	85.1	88.3	73.8	95.1	96.8
USA	76.9	91.9	56.6	86.7	94.1	93.3	98.5	96.3
Goodness-of-fit chisquared for the APC model (on 40 d.f.)								
Austria	2782.6	361.3	548.4	1417.5	1813.2	3173.5	2006.9	1397.5
Canada	5788.4	5930.3	607.1	1523.9	1997.9	1518.8	1586.0	1586.7
France	30775.3	10949.0	5930.7	13237.3	8537.6	8514.3	1884.2	3541.5
Hungary	27710.0	3944.2	1073.3	1483.1	14887.9	3579.6	8938.3	11921.8
Italy	16618.9	1857.9	1533.3	946.9	3167.3	1464.9	7245.9	10080.3
Japan	26701.9	2775.7	2460.1	3160.7	280472.3	184672.1	163166.1	42352.9
Poland	22969.9	3428.2	3598.6	741.1	257570.5	81279.5	28659.4	21764.5
Sweden	345.7	500.1	2299.4	3147.1	3384.3	1638.5	699.7	661.1
Switzerland	468.9	237.7	1144.5	618.4	1960.5	1376.4	309.6	802.7
UK	1347.8	2222.1	5023.7	2539.2	67362.0	71648.5	10106.3	6950.9
USA	27505.6	31470.6	16257.1	41587.7	171334.6	93527.0	8760.2	19165.2

[1] 100 (CHI_A - CHI_{APC}) / CHI_A where CHI_A is the goodness-of-fit chisquared statistic based on predictions using model A and CHI_{APC} is that based on predictions using model APC.

Table 6. Fitting the APC model to data for 1981-2010

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
Percentage of variance explained by period and cohort[1]								
Austria	98.4	95.5	87.8	81.4	99.5	98.2	99.8	99.6
Canada	99.6	99.2	98.1	87.8	99.5	99.8	99.3	99.5
France	97.3	99.4	99.2	98.4	99.8	99.8	99.9	99.9
Germany	99.0	99.6	99.6	85.5	99.9	99.8	99.9	99.8
Hungary	95.0	99.3	98.4	91.7	98.1	93.5	98.3	99.4
Italy	99.8	97.7	99.9	99.5	99.8	99.8	99.9	99.9
Japan	98.6	93.3	98.7	99.2	98.2	99.6	99.6	99.8
Poland	98.6	99.2	99.5	97.7	95.8	91.1	92.7	96.8
Sweden	94.5	98.7	92.4	95.6	99.9	99.7	98.7	98.6
Switzerland	98.5	97.2	98.8	70.9	99.7	98.9	98.9	99.4
UK	99.9	98.8	99.4	97.5	99.8	99.9	99.7	99.7
USA	99.7	99.4	98.8	99.5	99.6	99.8	98.9	99.4
Percentage of variance explained by cohort[2]								
Austria	91.0	89.3	52.9	67.3	92.3	70.7	80.2	85.7
Canada	98.1	98.0	68.4	81.6	69.6	82.7	45.6	42.6
France	94.3	96.5	49.0	61.2	94.2	95.8	88.5	95.4
Germany	93.0	97.8	88.5	74.7	98.9	97.0	97.1	94.3
Hungary	89.7	96.4	88.3	82.8	95.7	86.4	48.1	72.4
Italy	99.6	90.1	97.7	88.2	97.1	90.0	91.4	90.0
Japan	97.4	87.9	75.3	58.0	95.9	96.5	87.0	93.5
Poland	98.3	95.1	91.9	71.4	91.2	84.7	72.5	88.5
Sweden	90.0	93.5	82.3	92.8	89.1	74.7	58.0	50.9
Switzerland	79.4	81.3	89.2	64.3	93.8	65.4	48.4	80.9
UK	96.5	98.7	76.4	96.7	93.0	96.6	87.0	83.8
USA	98.8	99.2	91.9	97.7	79.9	95.4	75.5	61.8
Percentage variance explained by period[3]								
Austria	60.8	14.3	46.0	63.2	86.7	70.6	89.0	71.9
Canada	73.7	74.7	25.0	24.4	18.0	50.7	60.6	62.5
France	91.1	19.6	94.4	93.5	76.2	78.5	71.9	81.5
Germany	76.2	68.9	68.9	38.9	96.7	96.8	97.2	90.7
Hungary	89.6	84.0	89.9	61.1	65.6	24.9	80.9	82.2
Italy	75.0	25.8	80.7	78.5	62.9	63.4	22.4	9.5
Japan	90.7	76.3	50.5	70.9	91.0	93.3	85.7	83.0
Poland	74.0	35.1	91.4	81.8	67.5	54.1	85.0	90.2
Sweden	1.9	18.5	46.1	55.6	47.8	57.1	80.9	57.5
Switzerland	29.1	27.0	44.0	15.0	70.7	67.6	25.3	49.4
UK	85.3	84.8	33.3	14.0	87.2	96.3	76.8	66.0
USA	80.9	85.2	59.7	90.1	71.5	93.5	69.3	83.6

Table 6 continued

	Lung cancer		COPD		IHD		Stroke	
	Males	Females	Males	Females	Males	Females	Males	Females
Goodness-of-fit chisquared for the APC model (on 32 d.f.)[4]								
Austria	26.2	36.1	66.6	48.1	77.5	127.1	30.3	56.7
Canada	30.4	35.8	113.8	109.1	349.7	75.8	70.4	39.3
France	73.2	57.8	104.3	80.6	103.3	55.9	66.0	37.6
Germany	166.1	46.9	145.1	320.7	116.6	146.4	62.7	155.0
Hungary	85.2	26.5	32.9	66.6	115.2	181.1	164.7	96.5
Italy	41.5	46.2	45.4	32.2	136.4	88.4	76.8	100.0
Japan	72.4	62.1	243.1	109.3	322.5	102.4	680.3	1760.0
Poland	97.8	44.0	72.4	68.0	1313.5	1061.3	335.3	183.8
Sweden	33.6	27.9	57.4	37.2	43.8	46.2	41.9	45.1
Switzerland	37.2	29.4	28.6	54.5	38.8	53.7	51.7	23.6
UK	57.3	57.6	166.2	176.8	536.6	136.9	150.1	158.7
USA	199.1	221.3	96.0	135.2	2045.9	328.6	564.3	259.4

[1] 100 (RSS_A - RSS_{APC}) / RSS_A where RSS_A is the residual sum of squares for model A and RSS_{APC} is that for model APC. Critical values are 53.3, 60.1 and 67.2 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[2] 100 (RSS_{AP} - RSS_{APC}) / RSS_{AP} where RSS_{AP} is the residual sum of squares for model AP. Critical values are 45.4, 52.8 and 60.8 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[3] 100 (RSS_{AC} - RSS_{APC}) / RSS_{AC} where RSS_{AC} is the residual sum of squares for model AC. Critical values are 25.1, 33.2 and 43.10 for, respectively, $p<0.05$, $p<0.01$ and $p<0.001$.
[4] Critical values are 46.2, 53.5 and 62.5 for, respectively $p<0.05$, $p<0.01$ and $p<0.001$.

Table 7. Observed age standardized[1] death rates at age 30-79 in 1981-1985 and 1996-2000, and rates predicted by the APC[2] model in 2011-2015 and 2026-2030

	Males				Females			
	1981-1985	1996-2000	2011-2015	2026-2030	1981-1985	1996-2000	2011-2015	2026-2030
Lung cancer								
Austria	1129.0	925.7	665.1	430.4	185.2	253.4	335.4	339.9
Canada	1297.1	1078.7	687.3	376.5	378.3	583.9	566.7	402.0
France	1052.6	1108.0	910.4	628.5	93.7	161.2	312.8	554.2
Germany	1170.2	994.4	691.5	439.7	142.6	231.2	340.7	408.7
Hungary	1530.5	1956.7	1620.6	942.7	253.2	469.4	678.3	719.5
Italy	1369.9	1174.9	714.9	370.7	150.5	185.6	233.4	267.9
Japan	624.0	690.7	572.6	504.4	171.1	181.6	163.4	143.3
Poland	1483.1	1681.7	1266.4	667.2	180.5	282.3	417.1	426.6
Sweden	567.5	509.6	374.1	202.4	185.5	305.9	366.8	252.6
Switzerland	1183.5	828.7	566.6	340.6	141.2	229.0	329.4	325.5
UK	1582.4	974.4	646.0	441.5	458.2	483.7	496.6	523.5
USA	1337.2	1130.2	704.6	387.6	467.1	620.9	491.8	330.3
COPD								
Austria	455.6	327.2	261.9	157.4	137.2	102.1	125.7	90.6
Canada	517.5	373.2	216.8	144.4	163.2	213.4	177.5	124.7
France	337.9	265.8	103.4	62.4	96.5	92.9	36.9	24.9
Germany	673.3	416.6	241.1	175.2	161.4	143.8	134.8	130.1
Hungary	969.9	655.5	811.0	1260.0	297.2	228.5	342.5	591.7
Italy	583.7	298.8	142.3	86.0	141.4	79.7	54.1	38.9
Japan	228.3	151.0	65.7	35.0	86.9	47.2	17.2	8.3
Poland	833.8	463.7	330.9	226.9	192.3	113.9	107.6	91.6
Sweden	293.7	224.3	146.2	75.7	145.2	171.4	151.6	69.3
Switzerland	481.2	337.5	153.2	68.2	110.3	114.7	101.8	71.5
UK	875.1	555.5	328.2	222.7	292.0	356.2	267.4	189.3
USA	566.2	521.1	419.9	395.6	236.4	379.7	360.2	333.8
IHD								
Austria	3352.9	2403.5	1045.9	553.5	1168.5	874.7	342.3	155.2
Canada	3899.5	1978.2	950.1	519.4	1484.6	747.1	319.8	175.4
France	1533.0	945.1	450.0	224.0	494.0	255.6	109.3	71.3
Germany	3342.1	2193.5	953.3	471.0	1154.7	837.8	309.0	159.8
Hungary	4625.1	4205.4	3213.1	2088.5	1907.8	1819.1	1283.6	849.7
Italy	2165.5	1353.3	640.6	312.7	758.4	452.3	204.9	105.2
Japan	749.3	621.4	469.3	442.2	363.8	242.3	145.1	128.2
Poland	2624.2	2772.9	1447.9	543.4	765.8	947.4	453.6	160.0
Sweden	4416.6	2172.5	925.3	432.4	1468.3	752.6	345.7	197.1
Switzerland	2372.8	1483.2	686.9	397.9	719.7	481.1	190.9	101.5
UK	5118.9	2856.6	1037.9	439.9	1926.1	1120.3	347.7	147.2
USA	3943.7	2288.1	1224.4	686.7	1608.0	1012.7	503.0	279.8

Table 7 continued

	Males				Females			
	1981-1985	1996-2000	2011-2015	2026-2030	1981-1985	1996-2000	2011-2015	2026-2030
Stroke								
Austria	1651.1	798.2	230.0	81.9	1084.0	506.9	175.9	86.5
Canada	684.2	423.2	201.8	101.9	498.5	302.5	155.1	78.8
France	1018.3	494.1	230.6	114.9	584.6	270.0	138.5	83.6
Germany	1252.4	698.0	284.7	141.5	849.9	445.9	[18]5.8	107.3
Hungary	2895.6	2219.7	1019.4	377.9	1965.4	1290.5	485.5	158.7
Italy	1388.4	654.7	298.3	151.3	931.1	423.4	188.8	102.2
Japan	1860.5	915.2	501.1	308.9	1158.1	484.2	216.2	114.4
Poland	1023.7	1388.4	888.3	449.8	795.4	918.1	460.5	193.1
Sweden	820.1	625.0	286.6	126.2	598.8	403.9	194.3	84.3
Switzerland	790.1	384.2	173.2	90.1	524.0	255.6	137.4	89.6
UK	1215.3	693.2	290.0	129.3	958.8	544.8	227.4	100.9
USA	689.2	480.0	280.3	169.4	529.7	383.0	218.7	121.4

[1] Standardized to the 1976 European Standard Population.
[2] The APC model was fitted to data or 1981 to 2010 and used to predict data for 2011 to 2030 using Approach A for extrapolating period values and Approach 3 for extrapolating cohort values.

Table 8. Correlation[1] between cohort values and cigarette consumption per person at age 30-34 for that cohort

	Lung cancer		COPD		IHD		Stroke	
Country	Males	Females	Males	Females	Males	Females	Males	Females
Austria	0.76	0.56	0.85	0.16	0.85	-0.02	0.04	0.67
Canada	0.97	0.89	0.95	0.90	0.92	0.26	0.90	0.80
France	0.68	0.95	0.40	-0.36	0.84	0.23	0.92	-0.77
Germany	0.89	0.82	0.96	0.36	0.32	-0.34	0.12	0.57
Hungary	0.94	0.89	0.59	0.71	0.90	0.21	0.95	0.40
Italy	0.47	0.52	0.27	-0.86	0.37	-0.89	0.32	-0.85
Japan	0.77	-0.58	-0.03	-0.57	-0.48	0.80	-0.59	-0.63
Poland	0.68	0.73	0.60	-0.24	0.62	-0.55	0.81	-0.34
Sweden	0.86	0.82	0.77	0.36	0.58	-0.03	0.80	0.71
Switzerland	0.12	0.68	-0.01	0.95	-0.17	-0.21	-0.44	0.74
UK	0.92	0.82	0.83	0.96	0.90	-0.77	-0.65	0.58
USA	0.97	0.90	0.20	0.26	0.65	-0.60	0.62	0.04
Mean correlation[2]	0.75	0.67	0.53	0.22	0.52	-0.16	0.32	0.16

[1] Pearson correlation coefficient.
[2] Corresponding mean correlations for cigarette consumption at age 40-44 were 0.51, 0.64, 0.36, 0.14, 0.45, -0.19, 0.34 and -0.10. For age 50-54 they were 0.45, 0.57, 0.46, 0.23, 0.62, -0.29, 0.46 and -0.23.

10 References

1. Osmond C, Gardner MJ, Acheson ED. Analysis of trends in cancer mortality in England and Wales during 1951-80 separating changes associated with period of birth and period of death. *Br Med J* 1982;**284**:1005-8.

2. Osmond C, Gardner MJ. Age, period and cohort models applied to cancer mortality rates. *Stat Med* 1982;**1**:245-59.

3. Osmond C. Using age, period and cohort models to estimate future mortality rates. *Int J Epidemiol* 1985;**14**:124-9.

4. Peltonen M, Asplund K. Age-period-cohort effects on stroke mortality in Sweden 1969-1993 and forecasts up to the year 2003. *Stroke* 1996;**27**:1981-5.

5. Peltonen M, Asplund K. Age-period-cohort effects on ischaemic heart disease mortality in Sweden from 1969 to 1993, and forecasts up to 2003. *Eur Heart J* 1997;**18**:1307-12.

6. Marinaccio A, Montanaro F, Mastrantonio M, Uccelli R, Altavista P, Nesti M, *et al*. Predictions of mortality from pleural mesothelioma in Italy: a model based on asbestos consumption figures supports results from age-period-cohort models. *Int J Cancer* 2005;**115**:142-7.

7. Tzeng I-S, Lee W-C. Forecasting hepatocellular carcinoma mortality in Taiwan using an age-period-cohort model. *Asia Pac J Public Health* 2011;Epub ahead of print Oct 6:

8. Bashir SA, Esteve J. Projecting cancer incidence and mortality using Bayesian age-period-cohort models. *J Epidemiol Biostat* 2001;**6**:287-96.

9. Bray I, Brennan P, Boffetta P. Recent trends and future projections of lymphoid neoplasms--a Bayesian age-period-cohort analysis. *Cancer Causes Control* 2001;**12**:813-20.

10. Wong IOL, Cowling BJ, Schooling CM, Leung GM. Age-period-cohort projections of breast cancer incidence in a rapidly transitioning Chinese population. *Int J Cancer* 2007;**121**:1556-63.

11. Eilstein D, Uhry Z, Lim T-A, Bloch J. Lung cancer mortality in France: trend analysis and projection between 1975 and 2012, using a Bayesian age-period-cohort model. *Lung Cancer* 2008;**59**:282-90.

12. Cleries R, Martínez JM, Valls J, Pareja L, Esteban L, Gispert R, *et al*. Life expectancy and age-period-cohort effects: analysis and projections of mortality in Spain between 1977 and 2016. *Public Health* 2009;**123**:156-62.

13. Jürgens V, Ess S, Phuleria HC, Früh M, Schwenkglenks M, Frick H, *et al*. Tobacco-related cancer mortality: projections for different geographical regions in Switzerland. *Swiss Med Wkly* 2013;**143**:w13771.

14. World Health Organization. *WHO mortality database*. 2013, 1 May (downloaded 11 September). http://www.who.int/healthinfo/statistics/mortality_rawdata/en/index.html;http://www.who.int/healthinfo/statistics/mortality_rawdata/en/index1.html

15. *World population prospects: the 2012 revision. Excel tables - population data*. United Nations, Department of Economic and Social Affairs, Population Division, Population Estimates and Projections Section; 2013. http://esa.un.org/wpp/Excel-Data/population.htm;http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012_HIGHLIGHTS.pdf;http://www.un.org/en/development/desa/population/publications/pdf/trends/WPP2012_Wallchart.pdf

16. Forey BA, Hamling J, Lee PN. *International mortality and smoking statistics IMASS Version 4.09.* Sutton, Surrey: P N Lee Statistics and Computing Ltd; 2012. http://www.pnlee.co.uk/IMASS.htm
WARNING:There may be a more recent release version on the web site - this RefID created 4-Dec-2012 but pdf not updated

17. 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**:657-65.

18. Waterhouse J, Muir C, Correa P, Powell J, editors. *Cancer incidence in five continents,* Volume III. Lyon, France: International Agency for Research on Cancer; 1976. IARC Scientific Publications No. 15.