CANADIAN PARTNERSHIP AGAINST CANCER
PARTENARIAT CANADIEN CONTRE LE CANCER
CANCER SURVEILLANCE AND EPIDEMIOLOGY NETWORKS
RÉSEAUX DE SURVEILLANCE ET D'ÉPIDÉMIOLOGIE DU CANCER

Alberta Health Services

Cancer Projections Network (C-Proj)

Long-Term Projection Methods: Comparison of Age-Period-Cohort Model-Based Approaches

Prepared by

Alberta Health Services for

the

Canadian Partnership Against Cancer

December 2010

Document prepared by:

Zhenguo Qiu, PhD, Biostatistician
Zhichang Jiang, MSc, Surveillance Analyst
Mengzhe Wang, MSc, Analytical Team Manager
Juanita Hatcher, PhD, Director, Principle Investigator
And the Cancer Projection Analytical Network Working Team

Cancer Surveillance, Surveillance and Health Status Assessment, Population and Public Health, Alberta Health Services

Production of this document has been made possible through a financial contribution from Health Canada, through the Canadian Partnership Against Cancer.

The views expressed herein do not necessarily represent the views of the Canadian Partnership Against Cancer nor that of Health Canada.

Preferred Citation: Cancer Projections Network (C-Projections). Long-Term Projection Methods: Comparison of Age-Period-Cohort Model-Based Approaches. Alberta Health Services: 2010-12-xx

© Permission has been granted to the Canadian Partnership Against Cancer and Alberta Health Services to publish this document. The authors reserve other publication and distribution rights.

Highlights

Cancer incidence/mortality can be projected by extrapolating past trends using statistical models. The covariates used in these regression models are age, calendar period and/or birth cohort. Thus age-period-cohort analysis is appropriate for long-term (say, up to 25 years) projections of cancer incidence/mortality. Age-period-cohort models can be fitted using different approaches, including the generalized linear model (GLM), the generalized additive model (GAM) and the Bayesian model (BM), all of which have different advantages. In this validation analysis report, we evaluate the performance of the GLM Nordpred methods with different parameter settings, the GAM polynomial and natural spline methods, the Bayesian methods with different prior settings, and Hybrid age-period modeling. Using age-sex-specific cancer incidence and mortality data from 1975 – 1994 for Alberta (sex-specific population ≈ 1.6 million), we compared projections based on the above methods using data with the actual observed incidence and mortality from 1995 – 2008. Our findings were:

- Overall, the Nordpred, the Hybrid method and the Bayesian non-parametric smoothing on age, period and cohort method (Bray, 2002) outperformed other approaches;
- The GAM polynomial and natural spline methods performed well in some situations while they produced unusual results in other situations;
- Overall, the Nordpred with default parameter setting performed better than the Hybrid method; however, the Hybrid method performed better in some situations, for example, when cohort effect was absent for large cancer sites.

Based on the comparisons undertaken in this report we recommend:

1. The significance of period and cohort effects should be determined through age-period-cohort analysis prior to implementing any projection analysis.
2. If cohort effects are present, the default Nordpred method (NP2) should be used;
3. If cohort effects are not present,
    a. for common cancer sites the modified Hybrid method should be used.
    b. for small cancer sites, data aggregation is required to apply the Hybrid method; or the default Nordpred method (NP2) can be used.
4. When the number of cases/deaths is very small, the Bayesian method (BM1) can be used.

Key words and phrases: Age-period-cohort analysis; Age-standardized rates; Bayesian inference; Cancer incidence & mortality; Cramér-von-Mises test; Generalized linear models; Hybrid approach; Kolmogorov-Smirnov test; Lexis diagram; Markov chain Monte Carlo; Nordpred; Polynomial function; Projection & prediction; Relative bias; Natural spline; Statistical inference & computation; Trends; Validation analysis

List of Abbreviations

ACR	Alberta Cancer Registry
AC	age and cohort model
AD	age and drift model
AP	age and period model
APC	age, period and cohort model
ASR	age standardized rate
Ave	average method
BM1	Bayesian method – smoothing on age, period and cohort effects with removal of the linear trend of age effect
BM2	Bayesian method – smoothing on period and cohort effects with fixed age effects from age groups
BM3	Bayesian method – smoothing on age effects only assuming period and cohort random effects
CCS	Canadian Cancer Statistics
CVM	Cramér & von Mises
DIC	deviance information criterion
E	expected
GAM	generalized additive model
GLM	generalized linear model
HY	Hybrid method
KS	Kolmogorov & Smirnov
LM	linear model
MCMC	Markov chain Monte Carlo
MLE	maximum likelihood estimate
NHL	Non-Hodgkin lymphoma
NMSC	non-melanoma skin cancer
Nordpred	The R-package for APC model-based projection method developed in Nordic countries
NP1	Nordpred using log-link
NP2	Nordpred using "power 5" link
NP3	Nordpred using drift parameter over entire period
NSM	natural spline method
O	observed
PM	Polynomial regression method
RB	relative bias

List of Tables and Figures

Contents

1. Introduction

Health specialists, planners and policy makers need information on the future cancer burden in their jurisdiction to assist in planning and policy formulation. Accurate information allows them to prioritize prevention activities, to allocate health services, and to evaluate the impact of certain interventions or treatments. Thus accurate projections of the future cancer burden are essential. Cancer burden is usually measured by four indicators: incidence, mortality, prevalence and survival. Cancer incidence and mortality data are routinely recorded in cancer registries, and are the basic data for cancer surveillance. Prevalence and survival can be derived from these basic indicators. This report will only consider projections of cancer incidence and mortality.

Recently, many developed countries have provided long-term (≤ 25 years in the future) cancer incidence projections (e.g. Møller, Fekjær & Hakulinen *et al*, 2002; Møller, Fairley & Coupland *et al*, 2007). For example, the Canadian Strategy for Cancer Control aims to reduce the number of new cancer cases by 1.2 million over the next 30 years[a], and Europe Against Cancer[b] had proposed an ambitious programme aiming to reduce cancer mortality by 15% by the year 2000. To assess these goals, cancer incidence and mortality projections are essential (for examples, see Boyle, d'Onofrio & Maisonneuve *et al*, 2003; Sharp, Black & Muir *et al*, 1996). To be able to interpret incidence and mortality projection results, it is necessary to distinguish the changes due to a changing risk pattern over time and the changes due to the size and age structure of the population. The changes in projected rates reflect the changing pattern of risk factors, while the projected numbers of new cases or deaths depend not only on the changing rates but also on the projected population size and age structure. In this report our focus is on long-term projections of cancer incidence/mortality rates.

Currently in Canada there are five different methods used for long term projections, all based on statistical regression modeling of past trends in incidence/mortality (Stewart & Xie, 2009). If data

[a] Available at http://www.cancer.ca/Canada-wide/How%20you%20can%20help/Take%20action/Advocacy%20 what%20were%20doing/~/media/CCS/Canada%20wide/Files%20List/English%20files%20heading/pdf%20not%20in%20publications%20section/CSCC%20discussion%20paper%20-%20PDF_1404842209.ashx
[b] Available at http://www.epha.org/a/591

on risk factor trends and interventions are available, population-based modeling incorporating risk factor information may also be used. Risk factor data are not usually available in Canada, as risk factor surveillance is in its developing stages. Intervention data, usually related to screening data, are available in some jurisdictions; however, with the exception of breast and cervical cancer screening, screening programs have been implemented too recently to be able to determine the long term effects on incidence and mortality.

Statistical modeling of past trends allows projected cancer incidence and mortality trends to be estimated by extrapolating time trends from observed rates. The number of new cancer cases or deaths is calculated by applying the estimated rates to projected population numbers. Projections based on the extrapolation of trends in cancer incidence and mortality over time assume that trends in risk behavior will remain stable, no intervention or screening program will be initiated, and there is no change in diagnostic techniques. Trends in cancer incidence or mortality may be described as trends over age at diagnosis or death, year of diagnosis or death (period), and/or year of birth (cohort). Age is the most important time scale which affects cancer risk; it characterizes the cumulative exposure of the body to carcinogens over time. Period effects correspond to events that change incidence risk regardless of the age group and are usually due to a changing environment. Cohort effects involve risk factors that are shared by a specific generation as they age together and can be considered as "that which is due to early nurture" (Case, 1956). It is noted that the assumption of unchanged trends in rates is very strong and may not be realistic. The trend of observed rates reflects the unobserved trend in cancer risk. Usually, the trend can be classified as (overall or age-specific) period trend and/or cohort trend, which lead to two classes of models: age-period models and age-period-cohort models. Mathematically, the trend can be described as linear or non-linear and different statistical modeling techniques can be used including parametric, semi-parametric and non-parametric models. Because different statistical methods can result in different cancer projections, it may be difficult to determine which method is most appropriate. The choice of the appropriate statistical-model is fundamental to produce valid cancer projections.

In general, period effects may alter the cancer risk in both the short-term and the long-term, cohort effects on cancer risk would be more important over a long period than a short period. Thus in

general short term projections are based on age-period models, while long term projections are more likely to be based on age-period-cohort models. For short-term cancer mortality projection, Lee, Dean & Semenciw (2010) compared the Poisson regression method (Dyba & Hakulinen, 2000; adopted by Public Health Agency of Canada before), the SAS Proc Forecast (PF) method (Wingo, Landis & Parker *et al*, 1998), the state-space model (SSM) method (Tiwari, Ghosh & Jemal *et al*, 2004), the joinpoint method (Kim, Fay & Feuer *et al*, 2000) and the Nordpred method (Møller, Fekjær & Hakulinen *et al*, 2003). They found that the Nordpred method, combining with the age-drift-period model with a Power link function when the number of cases/deaths is small, performed the best. All methods for the short-term projections have not included cohort effects, except the Nordpred methods, which is an age-period-cohort model.

In long-term cancer incidence/mortality projections, age-period-cohort model-based approaches are widely used. For example, Nordpred is an R package for the modified version of the generalized linear modeling method (Osmond, 1985) and incorporates the idea of the smoothing method by degenerating the linear trend over period; the lexis diagram method, with a R package called "Epi", fits the non-linear functions of age-period-cohort effects by natural spline (Carstensen, 2007); the MIAMOD/PIAMOD software incorporates the non-linear functions of age-period-cohort effects by polynomial or cubic spline (De Angelis, De Angelis & Frova *et al*, 1994; Verdecchia, De Angelis & Capocaccia, 2002); and the Bayesian inference methods, with Markov chain Monte Carlo techniques, smooth the trend of age, period and cohort effects (Bray, 2002) or smooth the trend of age effects with randomization in period and cohort effects (Clèries, Ribes & Esteban *et al*, 2006). When cohort effects are absent, age-period model-based methods, e.g. the Hybrid method (Qiu, Jiang & Hatcher, 2010), could be utilized as an alternative. Other methods, such as the two-dimensional spline (Clements, Armstrong & Moolgavkar, 2005), functional data analysis approaches (Ramsay & Silverman, 1997, Erbas, Hyndman, Gertig, 2007; Yasmeen, Hyndman & Erbas, 2010), and multivariate vector autoregressive time series models proposed by Dr. Otterstatter (for example, see Xie, 2010), in which the cohort effects are not taken into consideration, may be suitable for long-term cancer projections as well, as the correlation between age and period may reflect the cohort effects.

This report will evaluate the age-period-cohort model-based methods of Nordpred, the polynomial regression, the natural spline and Bayesian analysis for long-term cancer incidence and mortality projections. In Section 2, we describe the cancer registry data and population data in Alberta which is used as the sample database for the validation analysis. Several age-period-cohort based approaches for long-term projections are introduced in Section 3 with special focus on the Nordpred method and Bayesian methods. In Section 4, we compare results obtained from those projection methods with observed data. We discuss those methods and results and recommend the best practice in Section 5.

2. Data Materials

Age- and sex-specific incident cancer counts for cases diagnosed between 1975 and 2008 and cancer mortality counts for the same period were obtained from the Alberta Cancer Registry (ACR) by cancer type as defined in Table 1. The coding for all cancer cases was converted into ICDO3 to eliminate the effect of coding changes on trends and projections. The selection of cancer sites for incidence and mortality was based on the Report on Canadian Cancer Statistics (CCS) 2009 (Canadian Cancer Society, Statistics Canada, Provincial/Territorial Cancer Registries & Public Health Agency of Canada, 2009 Tables 2.3, 2.5) for specified sex and incidence/mortality "All Cancers" excluded non-melanoma of skin (NMSC), and "All Other Cancers" includes all those cancer sites not explicitly showed in the main tables (except NMSC). The analysis included 36 sex-specific sites for incidence and 34 sex-specific sites for mortality

Populations for 1975 - 2008 of Alberta were obtained from Statistical Canada. In Alberta, the male population increased from 920,883 in 1975 to 1,624,200 in 2008, and the female population from 887,807 to 1,601,200 over the same time period. The 1991 Canadian standard population is used to calculate age-standardized rate.

In the validation analysis, data from 1975 to 1994 was used as the projection base to estimate the projected cancer incidence/mortality for 1995-2008. These estimates were compared with the observed incidence/mortality for those years. The data were tabulated into 18 five-year age groups (0 - 4, 5 - 9, …, 80 - 84, ≥85), 4 five-year aggregated periods for the projection base (1975 - 1979, 1980 - 1984, 1985 - 1989, 1990 - 1994), and 21 overlapping 10-year birth-cohorts, identified by central year of birth (1890, 1895, …, 1900) from 1890 to 1990 (see Table 2). Essentially, the relationship, cohort = period – age, can be extended to 5-year aggregated age-groups and periods, and cohorts can be coded by $c = A - a + p$, where $a = 1, 2, \ldots, A$ is the index number of age groups, $p = 1, 2, \ldots, P$ is the index number of periods, and $c = 1, 2, \ldots, C$ is the index number of cohorts. And $C = A + P - 1$. It should be noted that there are fewer observations for the earliest and most recent cohorts which may lead to less precision in the estimates of these cohorts.

Table 1: Cancer types by sex and incidence/mortality

Cancer Type	Sex	Incidence [a]	Mortality
Prostate	M	C61.9	C61
Breast	F	C50	C50
Lung	M, F	C34	C34
Colorectal	M, F	C18-C21,C26.0	C18-C21,C26.0
Non-Hodgkin Lymphoma	M, F	Type 9590-9596,9670-9719, 9727-9729	C82-C85,C96.3
Melanoma	M, F	C44 (Type 8720-8790)	C43
Leukemia	M, F	Type 9733,9742,9800-9801,9805, 9820,9826,9831-9837,9840,9860-9861,9863,9866-9867,9870-9876, 9891,9895-9897,9910,9920,9930-9931,9940,9945-9946,9948, 9963-9964, 9823 and 9827, sites C42.0,.1,.4	C91-C95, C90.1
Kidney	M, F	C64.9, C65.9	C64-C65
Pancreas	M, F	C25	C25
Bladder (+ in situ)	M, F	C67	C67
Oral	M, F	C00-C14	C00-C14
Stomach	M, F	C16	C16
Brain	M, F	C70-C72	C70-C72
Liver	M	C22.0	C22.0,C22.2-C22.7
Thyroid	F	C73.9	NA
Multiple Myeloma	M	Type 9731,9732,9734	C90.0, C90.2
Body of Uterus	F	C54-C55	C54-C55
Esophagus	M	C15	C15
Ovary	F	C56.9	C56
Testis	M	C62	NA
Cervix	F	C53	C53
All Other Cancers[b]	M, F	All sites excluding above	excluding above
All Cancers[b]	M, F	All invasive sites	All invasive sites

[a] Coded by ICDO-3: the Third Edition of the International Classification of Diseases for Oncology, 2000.
[b] Excluding non-melanoma of skin for incidence.

Table 2: Re-coding for birth cohorts

Birth Cohort	Age group																	
	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
-1890																		1
1895																	1	2
1900																1	2	3
1905															1	2	3	4
1910														1	2	3	4	
1915													1	2	3	4		
1920												1	2	3	4			
1925											1	2	3	4				
1930										1	2	3	4					
1935									1	2	3	4						
1940								1	2	3	4							
1945							1	2	3	4								
1950						1	2	3	4									
1955					1	2	3	4										
1960				1	2	3	4											
1965			1	2	3	4												
1970		1	2	3	4													
1975	1	2	3	4														
1980	2	3	4															
1985	3	4																
1990	4																	

Periods: 1=1975-1979, 2=1980-1984, 3=1985-1989, 4=1990-1994;
Age groups: 0 = 0 - 4, 5 = 5 - 9, 10 = 10 - 15, …, 80 = 80 - 84, 85 = 85^{+};
Birth cohorts: the central year of the overlapping 10-year birth-cohorts, e.g. 1895 = 1890 – 1899, 1950 = 1945 – 1954, etc.

Data formats may be different according to the requirements of Nordpred, Lexis diagram and Bayesian methods (see Section 3). For example, in the Lexis diagram, the 10-year overlapped cohort is split into two 5-year cohorts: one for the upper triangle (before the middle year), the other for lower triangle (after the middle year). Cancer cases can be tracked for the two cohorts according to the birth date of cancer patients in cancer registry database, and the approximate age-specific population size can be calculated use the integration formula proposed by Carstensen (2007).

3. Age-Period-Cohort Methods for Long-Term Projections

When long-term projections are of interest, both period and birth-cohort effects should be considered to be included in statistical modeling. This forms the age-period-cohort model. Age-period-cohort models can be fitted using different approaches, such as generalized linear models (GLM) (McCullagh & Nelder, 1989), generalized additive models (GAM) (Hastie & Tibshirani, 1990; Wood, 2006) and Bayesian models (BM).

Osmond (1987) first estimated future mortality rates using the age-period-cohort model under the framework of GLM. The age-specific observed death numbers were modeled by Poisson regression with age, period and cohort factors as regressors, and the logarithm of person-years as offset. Then the estimated age effects were the baseline (intercepts) to estimate the future age-specific rates adjusted by extrapolating the period- and cohort trends:

$$r_{ap} = \exp(\hat{\alpha}_a + \tilde{\beta}_p + \tilde{\gamma}_c), \quad c = A - a + p. \tag{1}$$

where A= number of age-groups, a: the a^{th} age-group, p= p^{th} future periods. The age effects $\hat{\alpha}_a$ were estimated from the GLM for historical data, and the fitted model also provided the period effects: $\hat{\beta}_p$ and cohort effects: $\hat{\gamma}_c$. Osmond (1987) modeled the linear trends for estimated period and cohort effects, respectively, and extrapolated the trends into the future to obtain the projected period effect $\tilde{\beta}_p$ and cohort effect $\tilde{\gamma}_c$ in the future p^{th} period. The computation can be implemented in many statistical software packages such as GLIM, SAS, and Splus / R etc.

The classical APC model sometimes gives unusual projections and thus has rarely been used recently (e.g. Shibuya, Inoue & Lopez, 2005; Clèries, Ribes, & Esteban *et al*, 2006). The two main issues for that method are: (i) whether the linear trends of the period effects and the cohort effects are realistic and (ii) whether too many parameters have to be estimated. If the projection is based on many insignificant estimated parameters, the predicted confidence interval will be very wide and thus will dilute the confidence on the projection.

3.1. Age-Period-Cohort Analysis

Age-Period-Cohort (APC) analysis can be used to discriminate statistically between trends due to period effects and trends due to cohort effects (Breslow & Day 1987). Analysis of historical data by age-period-cohort analysis can provide reliable estimated trends for projections. In this analysis, the retrospective 10-year overlapping birth cohorts are defined from the available data, although this has an obvious disadvantage, in that the cohort estimates for the more recent cohorts are based on fewer observations than the other cohorts (see Table 2). The data are generally grouped into 5-year periods and age groups, however, the corresponding cohorts are overlapping within adjacent cohorts (see Table 2). In age-period-cohort analysis, there is the non-identifiable problem as age, period and cohort are linearly related. Different methods had been proposed to address this issue, for example, reference period and cohort method (Decarli & La Vecchia, 1987); residual method (Holford, 1983), sequential method or spline method (Carstensen, 2007).

Recently, one-year period and age groups (and thus one-year birth cohort) have also been considered and age, period and cohort effects are regarded as continuous variables so that the polynomial regression or spline methods (such as B-spline, C-spline and natural spline) can be applied. However, one-year age-period-cohort analysis may not be feasible as the yearly age grouped population data (including historical and projected population) are not available in many provinces in Canada.

3.1.1. Preliminary Analysis:

In the preliminary examination of the data, four graphs were produced. They are (i) rates *vs* age, observations within each period connected, illustrating the cross-sectional age-specific rates, (ii) rates *vs* age, observations within each cohort connected, illustrating the longitudinal age-specific rates, (iii) rates *vs* period, observations within each age-group connected, illustrating the age-specific period rates and (iv) rates *vs* cohort, observations within each age-group connected, illustrating the age-specific cohort rates. Cohort effects are observed whenever there are different trends in age-specific rates over a range of cohorts; age effects are observed if age-specific rates vary by age for the same birth cohort or period. The period effects are observed if rates for all age groups change by period. For example, Figure 1 gives the four plots using Alberta lung cancer

incidence data 1972 - 2006, male, age 30 - 80+. Panel (i) and (ii) show the increasing rates by age-group; Panel (iii) shows the changing period effects for each age-group; and Panel (iv) shows the trends of cohort effects (rate at the logarithmic scale for better vision in young groups).

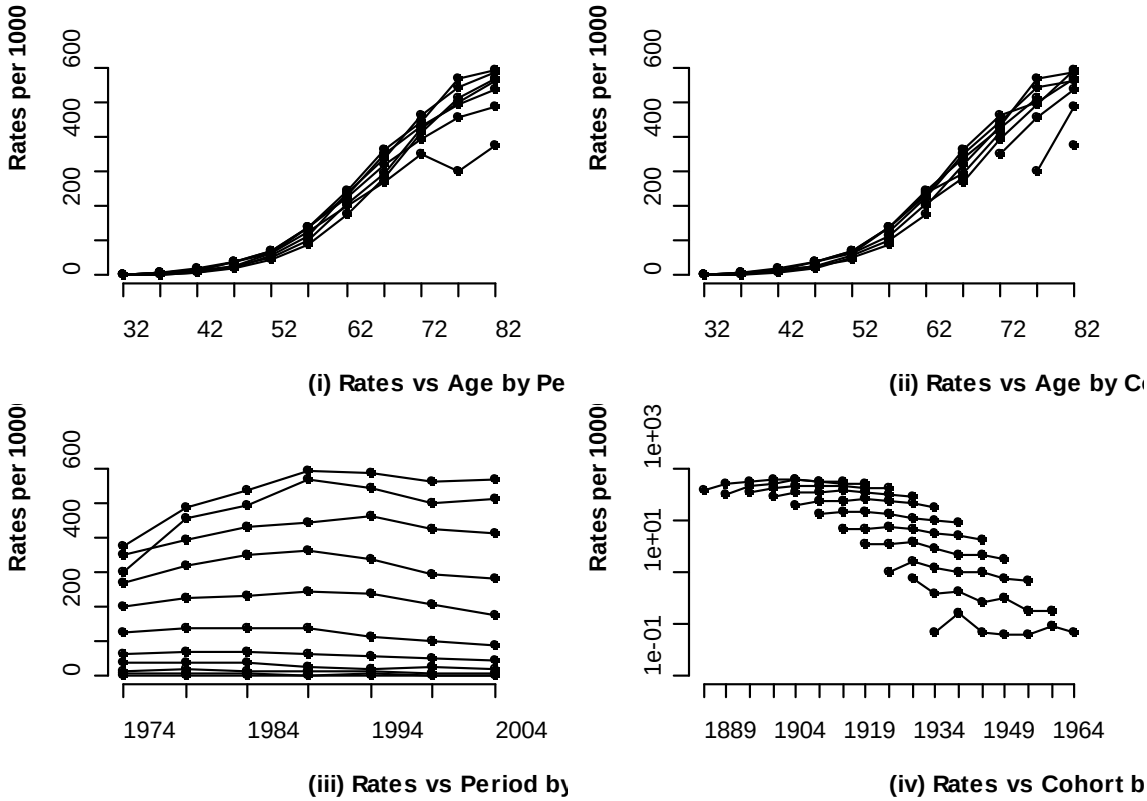

Figure 1. Preliminary age-period-cohort effects: Lung cancer incidence, Male, Alberta 1972-2006

3.1.2. Model Selection:

In general, log-linear modeling is used to estimate the age, period and cohort effects where the age, period and cohort are regarded as categorical variables. The model assumes that the number of incident cases in a specific age group and period is a random variable with a Poisson distribution. Under the framework of generalized linear model, the logarithm of the mean parameter was linearly linked to the explanatory variables: age, period, cohort and/or drift (i.e. the slope parameter of period or cohort when period or cohort is regarded as continuous variable, noting that the period drift and the cohort drift are equivalent due to the linear relationship among age, period and cohort), with the offset as the logarithm of the number of person-years (Clayton & Schifflers, 1987a, 1987b). The models were fitted using a stepwise approach. Age, age-drift, age-period, age-

cohort, period-cohort, and the full APC model are fitted to determine the best model parameterizations. Assuming the number of cases/deaths in age group a and period p: c_{ap} ~ Poisson(μ_{ap}) in person-year n_{ap}, the log-linear models for the above parameter combinations are as follows (noting that the drift from periods is equivalent to the drift from cohorts):

i. Age (A): $\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a$;

ii. Age-drift (AD): $\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a + D \times p$;

iii. Age-period (AP): $\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a + \beta_p$;

iv. Age-cohort (AC): $\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a + \gamma_c$;

v. Period-cohort (PC): $\log(\mu_{ap}) = \log(n_{ap}) + \beta_p + \gamma_c$; and

vi. Age-period-cohort (APC): $\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a + \beta_p + \gamma_c$.

Note that the period parameters β_p and the cohort parameters γ_c are the second order or higher (i.e. non linear effects). The drift parameter D is the linear trend. The goodness-of-fit of the models was evaluated by the deviance method. Comparison between the nested models was evaluated by the changes in deviance. Deviance changes between two models were assumed to be χ^2 distributed with the degree of freedom equal to the difference between the numbers of parameters in the two models. In most statistical software, ANOVA function with χ^2-test can be used to implement the selection procedure for nested models. However, the models in i – vi may not be nested at some level. The testing procedure should be (1) A *vs* AD, (2) AD *vs* AC and AD *vs* AP, and (3) AP *vs* APC, AC *vs* APC and PC *vs* APC. Particularly, we should pay attention to the comparison of the nested models A *vs* AD *vs* AP *vs* APC; If AP is the best, AP and AC should be tested to determine which variable (period or cohort) explains the historical data better. For the model if the ratio between the deviance and its degrees of freedom close to 1, the fit was considered adequate. If the ratio is far less than 1, the model over-fitted the data and a reduced model would be suggested. If the ratio is far greater than 1 for the model, the data are over-dispersed under the Poisson assumption. In such situation, the negative-binomial distribution can be invoked.

Under the nonparametric framework, for example, GAM, parameters in above models (excluding drift model) are changed to (non-linear) functions of those time variables (as continuous variables), that is, $f(a)$ substitutes α_a, $g(p)$ substitutes β_p, and $h(c)$ substitutes γ_c in above i – vi. Smoothing methods such as B-spline, cubic spline, natural spline and polynomial functions, or Bayesian non-parametric smoothing can be used to estimate the nonlinear functions.

Good projections are based on the appropriate analysis of historical data. Sometimes a simple regression model (Bray & Møller, 2006) can produce reasonable projections even for long-term objective. For example, when period and cohort effects are absent from APC analysis result, the average method may be the best way to estimate projections; when the cohort effect is not statistically significant, the age-period model methods can be the choice, e.g. the Hybrid method (HY) (Qiu, Jiang & Hatcher, 2010) could be used for long-term projection, with degenerated parameters on the period trend to dilute the impact of the historical trend over time.

3.2. Modified Generalized Linear Models — Nordpred

Based on the classic GLM method (Osmond, 1987), Møller (2004) proposed an approximate approach to estimate cancer incidence/mortality projections, and provided the R-package Nordpred for cancer incidence projections. The Nordpred method has been shown empirically to improve the projections (Møller, Fekjær & Hakulinen *et al*, 2003) with applications in European countries (e.g. Møller, Fekjær & Hakulinen *et al*, 2002; Møller, Fairley & Coupland *et al*, 2007).

In Nordpred, the "power 5" link function can be used to replace the log link function. The age-specific rate at period p, r_{ap}, can be written as

$$r_{ap} = (\alpha_a + D \times p + \beta_p + \gamma_c)^5 \qquad (2)$$

or

$$r_{ap}^{0.2} = (\alpha_a + D \times p + \beta_p + \gamma_c), \qquad (3)$$

where D is the common drift parameter, α, β and γ are the non-linear components of age a, period p and cohort c, respectively. Usually, 5-year aggregated data is used in Nordpred. One may use one-

year data in Nordpred with some revisions for the R code; however, this means that many parameters have to be estimated in the GLM and many estimated parameters may be insignificant.

3.2.1. Power *vs* Log Link Functions:

A problem with using the log link function in age-period-cohort modeling is that it produces projection that grow exponentially over time, which for some cancer types might give unrealistically high projections. In Nordpred, one can chose the "power" link functional form $h(x) = x^{0.2}$ instead of $h(x) = \log(x)$, and thus the projected rates increase as a function of x^5, which is less rapid than the exponential growth rates. Møller, Fekjær & Hakulinen *et al* (2003) found that the "power" link function improved projections of cancer incidence for the Nordic countries. The "power" link function, which is not the canonical link function for log-linear model, is an empirical selection from Box-Cox transformation theory, in which $\lim_{\rho \to 0} h(x) = x^{\rho}$ leads to the log-link function. The "power" link (ρ = 0.2) is situated mathematically between the log-link and an additive model (ρ = 1) (Møller, 2004).

3.2.2. Recent *vs* Over-All Period Trends:

For age-period-cohort modeling in 5-year periods, data for the most recent 20 years at least should be available. The trend (the drift parameter) in the most recent 10 years is generally suggested to capture the most recent linear trend. In Nordpred, the choice between using the trend over entire periods or a recent trend is based on the model:

$$r_{ap} = (\alpha_a + S_0 \times p + S \times p^2 + \gamma_c)^5 . \qquad (4)$$

If the estimated $\hat{S}$ was significant, the recent trend $D_{\text{last}} = D - \beta_P$ is used, otherwise D is used.

3.2.3. Cut-trend Parameters:

Due to the linear dependence among the three time variables: age, period and cohort, restrictions must be placed on parameters. In the full Nordpred model, the first (β_1) and the last non-linear period effects (β_P) are restrained to be equal, and the first (γ_1) and the last non-linear cohort effects (γ_C) equal. Any net period- and cohort-specific linear change is accounted for by a common component: the drift parameter D in (2) or (3) (Clayton & Schifflers, 1987b).

As suggested in Nordpred, the drift parameter (D or D_{last}) is attenuated gradually from 100%, to 75%, 50%, 25% and 0% in the future first, second, third, fourth and fifth 5-year period of the projection period, respectively (Møller, Fekjær & Hakulinen *et al*, 2002). This setting was used in this report. The future non-linear period (β_p) and cohort effects (γ_c) are assumed equal to the last estimated effects (β_P and γ_C) in the model (2) or (3).

In young age groups, where few cases occur, as is the case for most cancer sites, the age-specific trends would not be stable enough to provide significant trend estimation. In the Nordpred model the age specific rates for these age groups are estimated by the average method (Ave) based on the two most recent periods (10 years). In our analysis the age groups where this applies was set as those where the average number of new cases over the 5-year period was 5 for incidence (averagely at least one case per year) and the average number of deaths was 3 for mortality.

The age-period-cohort projection methods usually provide the estimated rates and numbers in five yearly periods. A linear interpolation method can be used to expand the estimated age-specific rates into annual rates. That is, for age-group a, the t^{th} year rate between the p^{th} and the $(p + 1)^{th}$ period is

$$r_{at} = \frac{5-t}{5} r_{ap} + \frac{t}{5} r_{a(p+1)}, \quad t = 1, 2, 3, 4, \tag{5}$$

where r_{ap} and $r_{a(p+1)}$ are the age-specific rates at age-group a in period p and $p + 1$, respectively.

In this validation analysis, we implemented three versions of the Nordpred projection model:

1. using the log-link function and the choice of drift parameter D or D_{last} based on testing (NP1),
2. the "power" link function and the choice of drift parameter D or D_{last} based on testing (NP2), and
3. the "power" link function with the fixed drift parameter D (NP3).

3.3. Generalized Additive Models

Another approach to age-period-cohort modeling is to consider the age, period and cohort variables as continuous, and use smoothing methods to fit that data. Two smoothing approaches will be investigated in this report:

1. smoothing using a polynomial function (PM), and
2. (natural) spline method (NSM).

These methods are under the framework of generalized additive models (GAM) (Hastie & Tibshirani, 1990; Wood, 2006), and may be particularly useful when annual data are available and the GLM would require an excess of parameters to be estimated. These methods can be also applied to the 5-year aggregated data; particularly, when the annual number of cases for single years of age is very small or even zero, leading to unstable rates, or the annual population data is not available by single years of age.

3.3.1. Polynomial Function:

In the polynomial method, age, period and cohort trends can be modeled by the log-linear regression for the mean parameter of the Poisson distribution μ_{ap}:

$$\log(\mu_{ap}) = \mu_0 + \sum_{i=1}^{I} \alpha_i a^i + \sum_{j=2}^{J} \beta_j p^j + \sum_{k=1}^{K} \gamma_k (p-a)^k \qquad (6)$$

where $\mu_0, \alpha_1, \cdots, \alpha_I, \beta_1, \cdots, \beta_J, \gamma_1, ..., \gamma_K$ are the parameters to be estimated. The linear term of period (i.e. $\beta_1 p$) was excluded in order to avoid the colinearity problem due to the linear relationship between age, period and cohort ($c = p - a$). The stepwise procedure can be used to select the best fitted model based on χ^2-test to decide the inclusion of the polynomial terms. The polynomial method can be implemented using the software MIAMOD / PIAMOD (De Angelis, De Angelis & Frova *et al* (1994) and Verdecchia, De Angelis & Capocaccia (2002)). This software can also fit the age-period-cohort model by the C-spline method. However, the polynomial expression is recommended by authors as in general it is hard to determine the degrees and turning points of the spline for inadequate time-points. It is noted that another version of the polynomial method was compared to the Nordpred method by Møller, Fekjær & Hakulinen *et al* (2003), in which the full polynomial model of Coleman, Esteve & Damiecki *et al* (1993) was reduced to

$$r_{ap} = \exp(\alpha_a + D \times p + \beta_2 p^2 + \beta_3 p^3 + \gamma_2 c^2 + \gamma_3 c^3) \qquad (7)$$

or

$$r_{ap} = \exp(\alpha_a + D \times p + \beta_2 p^2 + \gamma_2 c^2 + \gamma_3 c^3) \ . \qquad (8)$$

For 5-year age grouped data, this method may give better results; however, we have not repeated the comparison between this method and the Nordpred method.

Using the 5-year aggregated data in Alberta, we implemented the polynomial method (PM) in R. Linear interpolation (5) was used to produce the annual projections.

3.3.2. Natural Spline:

Carstensen (2007) proposed the natural spline to smooth the non-linear curves of the age-period-cohort model (yearly data by age, period and cohort was suggested) and extended the period and cohort effects into the future via non-linear interpolation for cancer rate projection. The general form is

$$\log(\mu_{ap}) = f(a) + g(p) + h(c), \qquad (9)$$

where *f*, *g* and *h* are non-linear spline functions for age, period and cohort effects, respectively. Sub-models (e.g., $f(a)$, $f(a)+g(p)$, $\log(\mu_{ap}) = f(a)+h(c)$ and $f(a)+D\times p$) may be selected from the age-period-cohort analysis using the spline method.

The approach can be applied using the R-package "Epi". We implemented the natural spline method (NSM) for the 5-year aggregated data in this report. It is noted that format of the age, period and cohort data is the triangular Lexis diagram and thus the double number of cohorts is obtained for both cancer data and population data.

3.4. Bayesian Methods

Bray (2002) had developed a Bayesian approach to smooth the age, period and cohort effects and estimate the age-specific rates from their posterior distribution by Markov chain Monte Carlo (MCMC) techniques (Robert & Casella, 2004). However, the use of Bayesian analysis, assumes knowledge of Bayesian statistics, in particular the recognition of the potential importance of prior distributions. It should also be noted that MCMC is inherently less robust than analytic statistical methods, and Bayesian computation is time consuming. For further information on Bayesian inference with Markov chain Monte Carlo techniques are given in Appendix B. The statistical inference and computation can be implemented in WinBUGS (Windows Version of Bayesian inference Using Gibbs Sampling) (Lunn, Thomas, Best & Spiegelhalter, 2000), or through other software such as R (BRugs or R2WinBUGS), Stata, SAS and Matlab (MATBUGS).

3.4.1. Bayesian Models:

Based on the log-linear model, in an age-period-cohort model under the Bayesian framework, the age-period specific number of cases or deaths, c_{ap} ~ Poisson (μ_{ap}), is defined as

$$\log(\mu_{ap}) = \log(n_{ap}) + \alpha_a + \beta_p + \gamma_c \tag{10}$$

where n_{ap} is the person-years at the p^{th} period and a^{th} age-group; $c = A + p - a$, with A=number of age groups (usually $A \leq 18$). The Bayesian method assumes priors for α_a, β_p and γ_c to assimilate and project the posterior probability distribution functions of the age, period and cohort effects, and the age-specific rates and counts as well, via Bayesian inference with MCMC computational techniques. We compared three different sets of priors for Bayesian analysis:

1. Transitional priors for age, period and cohort effects (BM1) (Bray, 2002),
2. Transitional priors for period and cohort effects and fixed age effects (BM2), and
3. Transitional priors for age effect only and random effects for period and cohort (BM3) (Clèries, Ribes & Esteban *et al*, 2006).

Details of prior distribution assumptions and hyper-prior assumptions are given in Appendix B.

3.4.2. Projections of Rates:

To project the future M periods, the simulated samples of fitted and projected rates are obtained through the combination of the simulated age, period and cohort samples by

$$r_{ap} = \exp(\alpha_a + \beta_p + \gamma_c),\ c = A + p - a,\ p = P + 1, P + 2, \ldots, M. \tag{11}$$

The projected age-specific number of cases can be calculated by multiplying age-specific rates with age-specific population size; or extending the model to future N period with "*NA*" (missing values) as the number of cases in data set, the estimated "missing value" ($\hat{\mu}_{ap}$) will be the age-specific number of cases in the future period p for the a^{th} age-group, noting that the "missing value" is regarded as parameter to be estimated in Bayesian analysis. In our experience, the median, instead of the mean, of samples is better to represent the "point" estimate of the parameter, including age, period and cohort effects, age-specific rates and standardized rate, and age-specific numbers and the total number of cases, as those parameters are usually skewed distributed. Outliers in a set of samples may distort the mean estimate. Figure 2 depicted the posterior density of projected ASRs of lung cancer incidence in the period 1995-1999, 2000-2004, 2005-2009 and 2010-2014 for female Albertans using BM1 method for data in 1975-1994. The densities illustrate an increasing

right-hand skewness along with the far future period. The difference between means and medians became larger (median [40.86, 47.01, 52.49, 57.73] *vs* mean [40.79, 47.38, 55.16, 67.43], in the 4 projection periods) and the 95% credible intervals became broader ([32.76-48.40], [29.32- 68.58], [24.29-98.87] and [19.67-146.0] in the 4 projection periods) as time goes on.

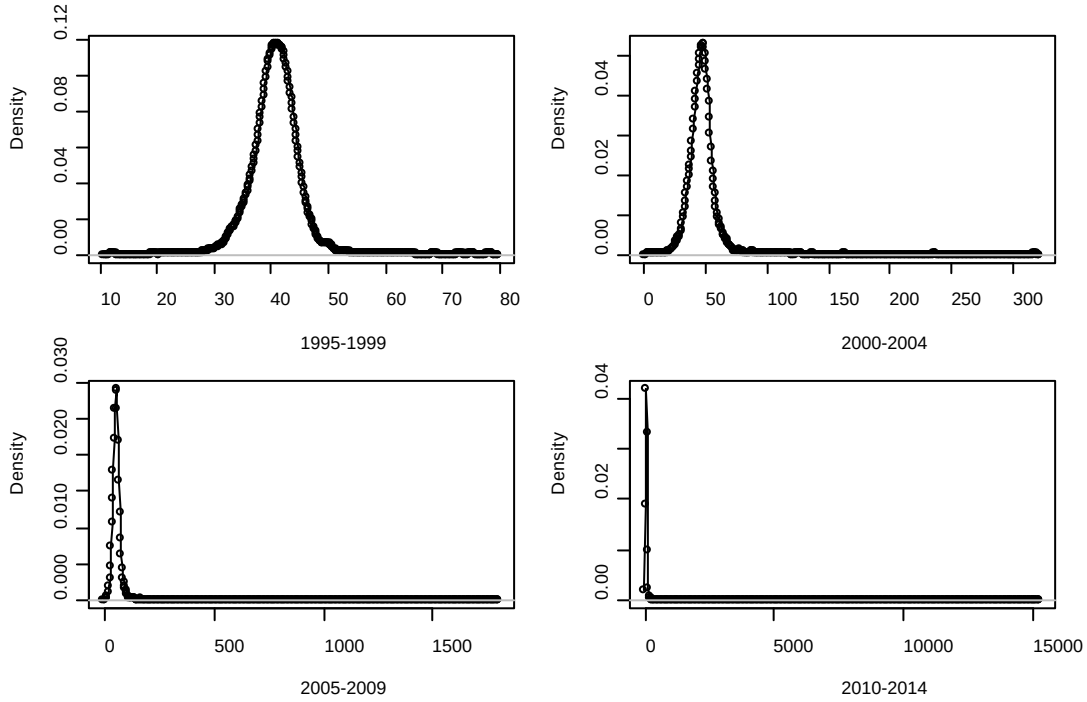

Figure 2. Posterior density of projected ASRs of lung cancer incidence in female Albertans

Any quantity related the age, period and cohort effects can be simulated in WinBUGS. For example, the annual projected rates can be obtained by linear interpolation (5) through the segment between the two mid-year points of two successive 5-year periods.

3.5 Comparisons of Methods

To compare the performance of these methods, the age-standardized incidence or mortality rates (ASR) are estimated for the years 1995-2008 by each different method using as the projection base data from 1975-1994. These projected rates were then compared to those observed in the same years. The relative bias was calculated for each year for which projections were calculated. The relative bias (*RB*) for the observed ASRs (*O*) and estimated ASRs (*E*) is defined as

$$RB_t = 100 \times \frac{|E_t - O_t|}{O_t}, \quad t = 1, 2, \ldots, 14. \tag{12}$$

The Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1948) was used to test overall performance for a given year. The other measure is the Cramér-von-Mises (CVM) statistic (Anderson, 1962; Xiao, Gordon & Yakovlev, 2007):

$$T = nw^2 = n \times \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \, dF^*(x), \tag{13}$$

where F^* is the empirical distribution of observed rates and F_n is the empirical distribution of projected rates. The CVM test was used to evaluate differences between the projected and observed rates over the 14 year of comparison base and to compare performance among methods for each cancer site by sex and incidence/mortality, The relative bias (*RB*) measures the absolute deviation (at the first moment) between observations and projections in 1995 and 2008, while the Cramér-von-Mises statistics (*T*) measures the sum of squared distance (at the 2nd moment) between the two empirical distributions of observations and projections The purpose was to determine the best method for each site by sex and incidence/mortality, and to compare the performance of each method among the 70 situations. Details of Kolmogorov-Smirnov (KS) test and CVM test are addressed in Appendix C.

The first step in the comparison was to identify the best method among the Nordpred models (NP1, NP2, NP3), GAM (PM, NSM), and the Bayesian models (BM1, BM2, BM3) separately. Then the performance of each of the preferred methods was compared, including the comparisons with the Hybrid model using the CVM test. Site specific projections were then compared among the models. In the comparisons, based on the observed and projected age standardized rates (ASRs) in 1995-2008, we provide

- Box plots for descriptive statistics of *RB*s to evaluate the performance of each method at the midpoint of each time interval (1997, 2002 and 2007) over 62 situations (The 8 situations with extremely large values from NSM or PM were removed, including prostate cancer incidence, to get better a vision of the plots);
- The *p*-values of KS test for *RB*s over the 70 situations to test the significant difference in performance for each pair of methods;

- Tables for CVM statistics (*T*) for projected ASRs and observed ASRs (14 points) by sex and cancer site incidence and mortality; the smallest *T* indicating the best projection;
- Box plot for CVM statistics over the 70 situations across the 9 projection methods;
- Tables of paired CVM test for *RB*s in 1995-2008 by sex-incidence/mortality-cancer type among HY, NP2 and BM1;
- Figures of the observed ASRs and projected ASRs by HY, NP2, NSM and BM1 by situation.

4. Results

4.1. Overall Comparisons within Methods

The overall comparisons of the variations within each method are detailed below:

i. *Nordpred*: Within the Nordpred methods, NP2 and NP3 are better than NP1 for longer-term (2002 and 2007) projections; this supports using the "power 5" link function for long-term projections (Figures 3 - 5); The difference between NP2 and NP3 is minor, as the use of the recent trend D_{last} is rarely indicated for only 20 years data; in some instances the NP3 model led to large *RB*s in the short-term projections; therefore, the NP2 is recommended; However, there are no significant differences among NP1, NP2, and NP3 (KS test *p*-value = 0.3291 for NP1 *vs* NP2, *p*-value = 0.1020 for NP1 *vs* NP3, *p*-value = 0.9995 for NP2 *vs* NP3). Comparing the methods using the Cramer-von-Mises statistic *T*, we see that within Nordpred methods, NP2 is slightly better than NP3, and although the median of the T statistic for the NP1 model was slightly smaller than that for the NP2 model, the interquartile range was larger for the NP1 model (Figure 6, and Table A3 and A4). In general, NP2 performed better than the other models.

ii. *GAM*: The NSM is better than PM method when eight situations with extremely large values, including the prostate cancer incidence, were removed from both the NSM and PM analysis; (KS test *p*-value < 0.0001); Both NSM and PM may produce unusual projections results. These results were confirmed by the Cramer-von-Mises statistic *T* values in Figure 6.

iii. *BM*: Within the Bayesian methods, BM1 and BM2 are much better than BM3 for the projections for 1997, 2002 and 2007 (Figures 3-5); (KS test, *p*-value < 0.0001 for BM1 *vs* BM3, and BM2 *vs* BM3). BM1 is slightly better than BM2 but not statistically significant (KS test *p*-value = 0.7842. Comparing using the Cramer-von –Mises statistic *T*, we see that BM1 is much better than BM3; and slightly better than BM2 (Figure 6, and Table A3 and A4). Thus BM1 performed better than the other models.

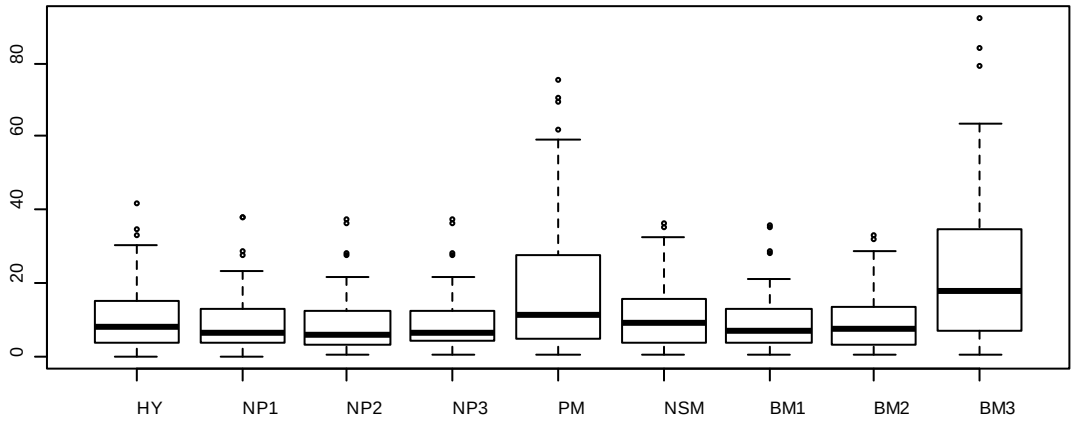

Figure 3. Box plot of relative bias in 1997, Alberta

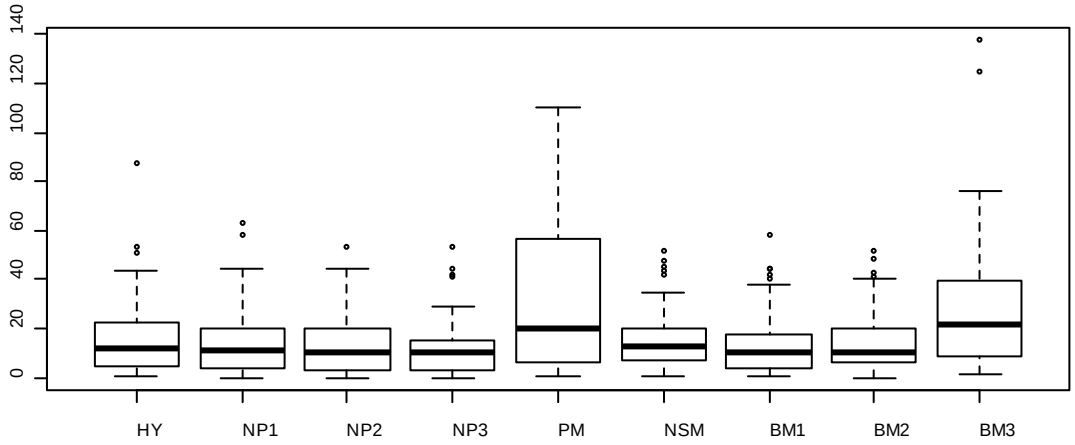

Figure 4. Box plot of relative bias in 2002, Alberta

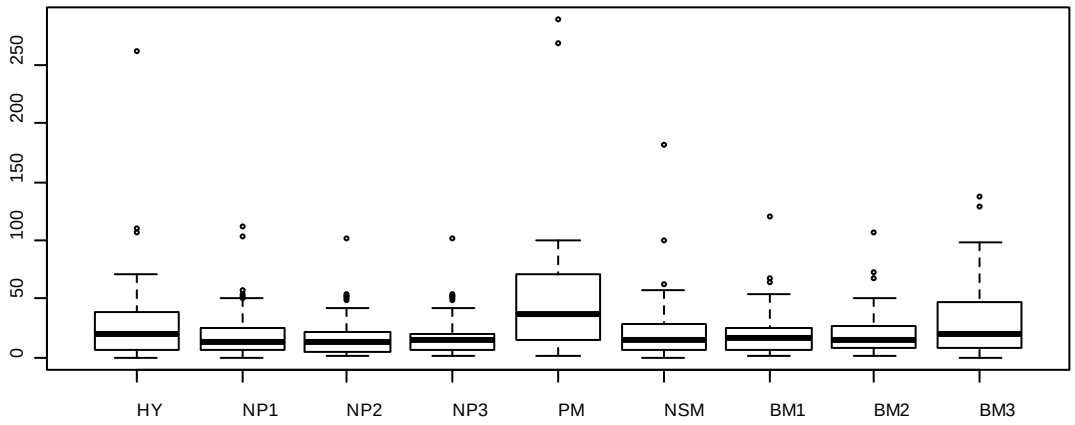

Figure 5. Box plot of relative bias in 2007, Alberta

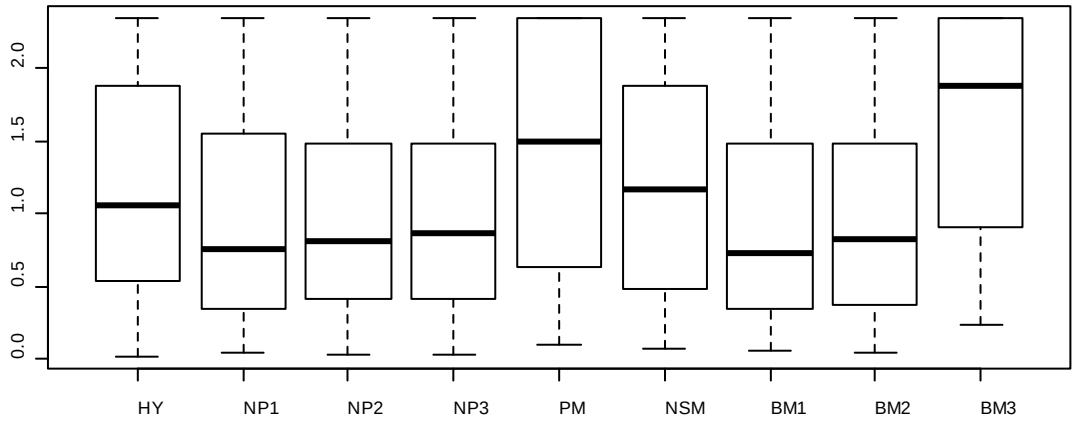

Figure 6. Box plot of Cramér-von-Mises statistics, Alberta

4.2. Overall Comparisons among Methods

The next step in the comparison was to compare the *RB*s among the best models of each type of model. These were HY, NP2, NSM and BM1. We found that (i) NP2 is better than NSM (KS-test p-value < 0.0001), and the HY method (KS-test p-value =0.0003), but not the BM1 method (KS-test p-value = 0.3021); (ii) BM1 is better than the HY method with KS-test p-value = 0.0101; (iii) the HY method is better than the NSM method KS-test p-value = 0.0348.

In the comparisons using Cramer-von-Mises statistics, the HY method is slightly better than NSM, but worse than the NP2 and the BM1 (Figure 6). It seems that the polynomial, the spline and the pure random effect assumptions of time trend are problematic for the trends of Alberta cancer data. Thus the GAM methods will not be discussed in the site specific comparisons.

4.3. Site Specific Comparisons

The results of the paired comparisons among HY, NP2 and BM1, using the CVM test based on the *RB*s for each sex-cancer site incidence/mortality, are presented in Tables A1 and A2. The CVM statistics T comparing the observed and projected ASRs obtained by each method for each sex-cancer site-incidence/mortality are presented in Table A3 and A4. The graphs of trends for each sex-cancer site-incidence/mortality including the projection results are presented in Figure A1 and A2.

4.3.1. BM1 *vs* NP2:

Of the 70 possible comparisons, BM1 gave smaller *T* values than NP2 in 32 situations in which 20 situations were for mortality projections, while the BM1 gave larger *T* values than NP2 in 29 situations in which 16 situations were for incidence projections. There were 9 ties between the *T* values from the BM1 and NP2 (see Table A3 and A4). BM1 is significantly better than NP2 in 3 situations (male and female all cancer mortality, and male NHL mortality), and the NP2 is significantly better than the BM1 in 4 situations (female lung cancer incidence, male melanoma and kidney incidence, and male multiple myeloma mortality). All other comparisons were not significant (Table A1 and A2),

In general, the performance of NP2 and BM1 is similar. The BM1 is slightly better than NP2 for mortality projections, while the NP2 is slight better than the BM1 for incidence projections. In general mortality rates are decreasing over the period 1995-2008, except for female lung, male NHL, male liver and all other cancers, while incidence is increasing for most cancer sites, except male lung, male pancreas, male oral, and ovary. Also, mortality numbers are usually smaller than the incidence numbers. It is noted that the BM1 used the exponential trend which is appropriate for decreasing trends, but the NP2 used the power 5-trend which is considered to be more appropriate for increasing trends. The Nordpred with log-link function (NP1) did not perform as well as the BM1 in mortality projection (13 (NP1) *vs* 19 (BM1), with 2 ties), but the NP1 performed better than the BM1 in incidence projection (19 (NP1) *vs* 9 (BM1), with 8 ties) and mostly in the situation with large numbers, e.g. female lung, breast, NHL, melanoma, leukemia, kidney, all other, and all cancers (see Table A3 and A4). As the Bayesian inference with the Markov chain Monte Carlo technique is more complex to implement, the BM1 should be used only for small number of cases/deaths as necessary.

4.3.2. HY *vs* BM1:

The BM1 gave smaller *T* values than the HY in 43 situations in which 22 situations were for mortality projections, while the BM1 gave larger *T* values than HY in 26 situations in which 14 situations were for incidence projections. There was only one tie between the *T* values from the

BM1 and HY (see Table A3 and A4). Except for the situation of cancer incidence of the body of uterus, the cohort effects were not significant in the age-period-cohort analyses for the 25 situations in which the HY outperformed the BM1.

BM1 is significantly better than HY in 14 situations, and HY is significant better than BM1 in 10 situations (female breast, colorectal, kidney and body of uterus cancer, and male colorectal cancer, leukemia and multiple myeloma incidence, and female colorectal cancer and melanoma, and male multiple myeloma mortality) (Table A1 and A2). In only five of the 14 situations, where BM1 was better than HY, the cohort effects were significant (male and female lung cancer incidence and mortality, and prostate cancer incidence); but for only one (body of uterus cancer incidence) out of the 10 situations, HY is better than BM1.

Body of uterus cancer incidence: All APC model-based approaches performed worse than the HY method (see Figure A1 and Table A3). The projection base for all APC model-based approaches is 1975-1994 and the high rates in 1975-1979 influenced the downward trend in the future. However, the HY method identified the projection base as 1980-1994 by the likelihood searching approach and used the negative binomial distribution-based log-linear model for age-specific rates (age $\geq$ 35). And thus HY obtained stable projected rates in 1995-2008 as observed. It is noted that the BM3 also performed well for this situation as the decreasing trend in 1975-1994 was not used for the projection.

Colorectal cancer incidence and mortality: The cohort effects were not significant for this common cancer type (Table A3 and A4). And thus HY outperformed BM1 for the projections of female incidence and mortality, and male incidence. Both HY and BM1 performed very well for male mortality (see Table A4).

Female breast cancer incidence: The cohort effects were not significant in this common cancer site. The age-period model-based approach (HY) gave better projections than those from the age-period-cohort model-based approaches (Table A3 and Figure A1).

Male leukemia incidence: The cohort effects were not significant, and the HY performed much better than BM1. Obviously the trend in the projection base of 1987-1994 used in HY is more relevant to the future trend than the trend in 1975-1994 or 1985-1994, noting the sharp change in 1985-1986 (Figure A1, and the Partnering Provinces Trend Report: Joinpoint Trend Analysis of Cancer Incidence and Mortality).

Female kidney cancer incidence: Again, the cohort effects were not significant and the HY performed much better than BM1. For kidney cancer, the trend based on 1985-1994 by HY is closer to the trend in 1994-2008 than the 2nd-order transition trend in the 4 period in 1975-1994 used in BM1 (Figure A1).

Female melanoma mortality: For melanoma of skin cancer mortality, the numbers were very small and thus the average method (from 1990-1994 data) was used in the HY model.

Male multiple myeloma incidence and mortality: Again, the cohort effects were not significant and the HY performed much better than BM1. The projection bases used were 1983-1994 for incidence and 1981-1994 for mortality in the HY. Therefore, the HY projected a downward trend in incidence rates, and stable rates in mortality, which are more similar to the trends observed. With cohort effects, the BM1 performed worse than the HY method.

4.3.3. NP2 *vs* HY:

In the Nordpred methods, each projection was estimated by extending the trend from the drift parameter. The drift parameter represents the linear trend in rates over time, and it is not possible to distinguish whether this is due to period effects or cohort effect. Thus it may represent linear trends over periods and/or cohorts. In this analysis, the NP2 method gave smaller *T* values than the HY method in 40 situations in which 19 situations were for mortality projections, while the NP2 method gave larger *T* values than the HY method in 28 situations in which 14 situations were for incidence projections. There were two ties between the *T* values from the NP2 and HY methods (see Table A3 and A4). In 26 of the 28 situations where the HY method gave smaller *T* values than the NP2 method, cohort effects were not statistically significant in the age-period-cohort analyses.

The two sites where cohort effects were significant in the model were body of uterus cancer incidence and female lung cancer mortality,

The NP2 method is significantly better than the HY method in 15 situations, and the HY method is significantly better than the NP2 method in four situations (female body of uterus and male colorectal cancer incidence, and female lung cancer and male NHL mortality) (Table A1 and A2,). The cohort effects were statistically significant for four situations (male lung cancer incidence and mortality, female lung cancer incidence, and prostate cancer incidence) out of the 15 situations (where the NP2 method was significantly better than the HY method). For two of the four situations where the HY method was significantly better than the NP2 method, cohort effects were statistically significant in the age-period-cohort analysis (body of uterus cancer incidence and female lung cancer mortality).

HY method better than NP2 method:

Body of uterus cancer incidence: All APC model-based approaches had higher *T* values than the HY method (see Figure A1 and Table A3). The projection base for all APC model-based approaches was 1975-1994 and the high rates in 1975-1979 led the downward trend in the future. However, the HY method identified the projection base as 1980-1994 by the likelihood searching approach and used the negative binomial distribution-based log-linear model for age-specific rates (age >= 35). And thus the HY method obtained stable projected rates in 1995-2008 as observed.

Female lung cancer mortality: Although the age-period-cohort model is recommended (Table A4), $T = 0.69$ from the HY method is smaller than the $T = 0.81$ from the NP2 method. However, other APC model-based methods, e.g. NP1, NP3, PM, NS, and BM1 obtained smaller *T* values than the HY method. In fact, the projection from the NP2 method is visually equivalent than that from the HY method (see Figure A2). We can say that the CVM test, which is based on the empirical distributions of *RB*s from the two methods, is not appropriate for this situation. In terms of the absolute relative biases (*RB*s) in 1995-2008, the NP2 method performed much better than the HY method, with the mean of *RB*s at 7.5 (*vs* 17.9 for HY) and median of *RB*s at 8.4 (*vs* 19.2 for HY).

Male colorectal cancer incidence: The age-period model was recommended for this common cancer site (Table A3). And thus the HY method outperformed the NP2 method in this situation. In the HY method, the negative-binomial distribution based log-linear model was used for age-specific rates (age ≥ 35) using the projection base of 1980-1994.

Male NHL mortality: The age-drift model was recommended for male NHL mortality. The rates are based on small numbers and are not stable, which may affect the choice of model. In the HY method, the Poisson distribution based log-linear model was used for common age trend (age ≥ 35) using the projection base of 1980-1994. And thus the HY method performed better than NP2. However, the NP3 and BM2 methods, which include cohort effects, performed even better than the HY method.

NP2 method better than HY method:

In the 15 situations in which the NP2 method is significant better than the HY method, the cohort effect was significant in the age-period-cohort analysis for only four situations. For the remaining nine situations (excluding male all cancer incidence and female all cancer mortality where it may be preferable to use the add-in approach), small numbers of cases or deaths were observed, and the age model was recommended for 5 situations and age-drift model was recommended for 4 situations from age-period-cohort analyses. We discuss the details as follows:

Female oral and stomach cancer incidence: The observed ASRs in 1975-1994 were less than 10 per 100,000. For oral cancer, the HY method identified the projection base as 1986-1994 for the annual data and used the Poisson distribution-based log-linear model with common age trend for age ≥ 40 (the average method was used for age < 40). The highest ASR was observed in 1986; this influenced the decreasing trend between 1986 and 1994 and biased the downward trend used in the projections (Figure A1). Similarly for stomach cancer, the projection base for the HY method was chosen as 1988-1994, where a slightly increasing trend was observed (Figure A1), but overall a decreasing trend was observed over the years in 1975-1994. Given the small number of cases, the annual data should be aggregated by certain period of year, e.g. 2 or 3 years, to smooth the rates,

before applying the HY method, which should provide more reliable trends with a re-defined projection base.

Male brain cancer incidence: In 1975-1994, the lowest ASR was observed in 1981. The HY method identified the projection base as 1981-1994 for the annual data and used the negative-binomial distribution-based log-linear model with age-specific trends in childhood (0-9) and ages greater than or equal to 25 (the average method was used for age 10-24). The smallest rates in the beginning point of the projection base led to an increasing trend and thus produced increasing trend in projections (Figure A1). Again, for the small cancer site (ASR < 9 per 100,000), the annual data should be aggregated.

Male testicular cancer incidence: The HY method identified the projection base as 1986-1994 for the annual data and used the Poisson distribution-based log-linear model with age common trend in 20-49 (the average method was used for the rest of age group). This fitted relatively stable rates in 1986-1994 and thus underestimated the projections in 1995-2008 (Figure A1). The annual data should be aggregated when applying the HY method for the small cancer site.

Female and male kidney cancer mortality: In 1975-1994, the highest ASR was observed in 1982 for male and the second highest ASR in 1987 for female (Figure A2). The HY method identified the projection base as 1982-1994 (male) and 1987-1994 (female) for the annual data and used the Poisson distribution-based log-linear model with age common trend (age ≥ 40 for male, and age ≥ 50 for female). Thus, projections were underestimated. Again, for those small numbers (ASR < 8 per 100,000), the annual data should be aggregated when applying the HY method.

Female ovary and cervix cancer mortality: The projection base was identified as 1983-1994 for ovary cancer and 1989-1994 for cervix cancer. The ASR is low in 1983 for ovary cancer and in 1989 for cervix and thus led the overestimation of projections (Figure A2). Again, the number is small (ASR < 10 for ovary cancer, and < 6 for cervix cancer), and data aggregation is required.

Male leukemia mortality: The ASRs were less than 12 per 100,000 as shown in Figure A2. The HY method identified the projection base as 1987-1994 for the annual data and used the Poisson distribution-based log-linear model with age common trend in age ≥ 35. An increasing trend was fitted for data in 1987-1994 and thus the projection was overestimated. However, over 1975-1994, the decreasing trend was observed (Figure A2). Considering the improvement of treatment for childhood leukemia, cohort effects can be incorporated for the projection.

5. Discussion and Recommendation

Cancer projections from statistical models should be used with caution. The projections rely on the assumption that past trends will continue into the future. This is a very strong assumption which is rarely justified in practice. For most cancer sites, however, the statistical modeling based approach is the only way to do cancer projections, as there is a lack of risk factor knowledge and data. Because of their uncertainty, projections from statistical models are usually just one part of the decision-making process.

Some cancer projections may change very much within a short period (refer to Figure A3), while other projected results may stay on the track for long time (refer to Figure A4). Therefore, the long-term projection (say, 20 years) does not mean that we should do the log-term projection after 20 years once the current projection works are done. Continuously upgrading projections is a typical task in cancer surveillance.

In this work, we implemented the Hybrid method, the Nordpred methods, the polynomial method, the natural spline method, and the Bayesian methods for Alberta 1975-1994 cancer data, and compared the projected results using the observed age-standardized rates in 1995-2008, based on the measurements of the relative bias and the Cramér-von-Mises statistic.

The methods evaluated in this report are all based on historical trends in rates. Future changes in rates cannot be forecast without knowledge of changes in risk behaviors, and screening practices for incidence and mortality and treatment for mortality. For example, the rapid increasing trend in 1996-2008 for female thyroid cancer incidence, and the increasing trend in 1990-2008 for male esophagus cancer incidence, can not be projected by any historical trend based methods. Similarly for prostate cancer, none of methods worked well as there have been significant changes in trends, probably related to the use of the PSA antigen test. Also, the recent decreasing trends in 1994-2008 for female breast cancer mortality and in 1995-2008 for prostate cancer mortality were not reflected in the projections. The decreasing trend in breast cancer mortality may reflect increased uptake of breast cancer screening, and projections based on the full range of data (1975-2008) would reflect these recent changes in trend, although they would not reflect any future changes.

Overall the polynomial regression method did not perform well although it may produce the best results in some situations (e.g. female brain cancer incidence, female bladder mortality, and male all cancer mortality). This may be due the form of the equation where inclusion of higher order powers (e.g. x^4) may lead to unusual results in some situations (e.g. male leukemia incidence and female kidney mortality). Similarly for the spline method, the non-linear trend may lead to unrealistic projections (e.g. male stomach incidence, male multiple myeloma incidence and mortality); while it performed very well in some instances (e.g. male lung cancer incidence and mortality in Figure A1 and A2). In general, these two methods may work well based on the yearly age, period and cohort data for large cancer sites in large population.

Bayesian age-period-cohort methods may be the choice for long-term projections for small cancer sites in small populations. Bayesian inference via MCMC has the advantage in dealing well with small numbers. For example, Maule *et al* (2006) adopted the Bayesian method, a random effect model for smoothing the time trends, proposed by Breslow & Clayton (1993), to detect the changes in the time trend for childhood cancer incidence. Over the three Bayesian methods (BM1, BM2 and BM3) used in this report, the BM3 (Clèries, Ribes & Esteban *et al*, 2006) performed least well. The BM3 adopted the random effects for both period and cohort effects and only used a transitional prior for the trend on age effects. This stabilizes the projected rates, but the existing historical time trend (in period and/or cohort) is not properly recognized. The BM2 method emphasizes the trend in both period and cohort by the 2^{nd} -order transitional prior, but the age effect is assumed to be fixed (corresponding to the categorical age effects in GLMs such as the Nordpred method). Zero cases can not be handled in this model. The removal of the age groups where this occurs may lead to under-estimated age-standardized rates in some situations. Although Bray's (2002) method (BM1) took longer computation time than BM2, we recommend the BM1 method in the sense of the capacity of dealing with small (even zero) number of cases or deaths. Though the BM1 method produced similar projections to the NP2 method in most situations (see Figure A1 and A2), we suggest that the Bayesian methods (BM1) are used only in small-number situations with significant cohort effects due to the complexity of Bayesian inference. In general if Bayesian methods are to

be used, we recommend the use of the BM1 method, but, when the modeled trend is not reliable, the BM3 method may give the best results.

In this analysis, the default setting of Nordpred slightly outperformed other two Nordpred methods. The "power 5" link plays an important role in long-term projection but the recent trend (D_{last}) must be considered carefully. It may be too sensitive to call the recent trend for small cancer sites. In the absence of the cohort effects in the age-period-cohort analysis, the Nordpred APC method may over-fit the data and thus provides worse projections than those from the HY methods (e.g. the colorectal incidence and mortality projections). In general, if estimated cohort effects are not significant this could be an issue for common cancers. When the values of the estimated cohort effects are very small (usually for rates based on small numbers), the impact on projections is minor. However the predicted confidence intervals may become wide due to the relative large standard error of the cohort estimates.

In general, the HY performed better than Nordpred if and only if the cohort effects were not significant from age-period-cohort analyses for the more common cancer sites. The HY approach includes the average method, the age-drift model, and the age-drift-period model based methods, using annual age-specific counts and population sizes, with identification of the projection base by likelihood searching method. For small numbers, e.g. small cancer sites, the identified cutting point of calendar year may not be reliable and thus provide inaccurate projections (e.g. the cervix cancer mortality projection in Figure A2, where the projection base was selected as 1989-1994). It is suggested that appropriate aggregation of data over years to increase the numbers in each period for small cancer sites. Interpolation functions can be used to convert the projected aggregated rates to annual rates. In addition, a different link functions such as the "power 5" should be used for long-term projections. Therefore, the Hybrid method should be modified for potential use in long-term projections, and further analysis is undertaken to compare the Hybrid model with the Norpred model for use with small cancer sites where there are no cohort effects. This further development of the Hybrid method may potentially provide an improved method for long term projections where there are no cohort effects. When the data are aggregated by 5-year, the revised Hybrid method would be similar to the age-drift-period model which was used for small population by Møller

(2004). The advantage of the Hybrid method is that it is easier to understand and apply than the Nordpred method. This further development is outside the scope of the current project.

For short term projections, PHAC has recommended the use of the full Nordpred model for common cancer sites, but for small cancer sites PHAC has recommended the use of a modified Nordpred model based on age-drift-period analysis rather than age-period-cohort analysis of the full model (Lee, Dean & Semenciw, 2010). However, our analysis has shown that for rates based on small number there may be significant cohort effects (e.g. body of the uterus cancer incidence), and similarly there may be no cohort effects for some of the more common cancer sites (e.g. colorectal and breast cancer in Table A1). This may affect the choice of projection methods.

We acknowledge that projections of cancer incidence and mortality based on historical trends may be inaccurate, whatever method used, if the underlying trend in risk changes. Thus it is important to review any cancer incidence or mortality projection on a regular basis, to determine if the underlying trend has remained constant.

Recommendation:

These recommendations are based on the comparisons undertaken in this report.

4. The significance of period and cohort effects should be determined through age-period-cohort analysis prior to implementing any projection analysis.
5. If cohort effects are present, the default Nordpred method (NP2) should be used;
6. If cohort effects are not present,
    a. for common cancer sites the modified Hybrid method should be used.
    b. for small cancer sites, data aggregation is required to apply the Hybrid method; or the default Nordpred method (NP2) can be used.
4. When the number of cases/deaths is very small, the Bayesian method (BM1) can be used.

Appendix A. Figures and Tables

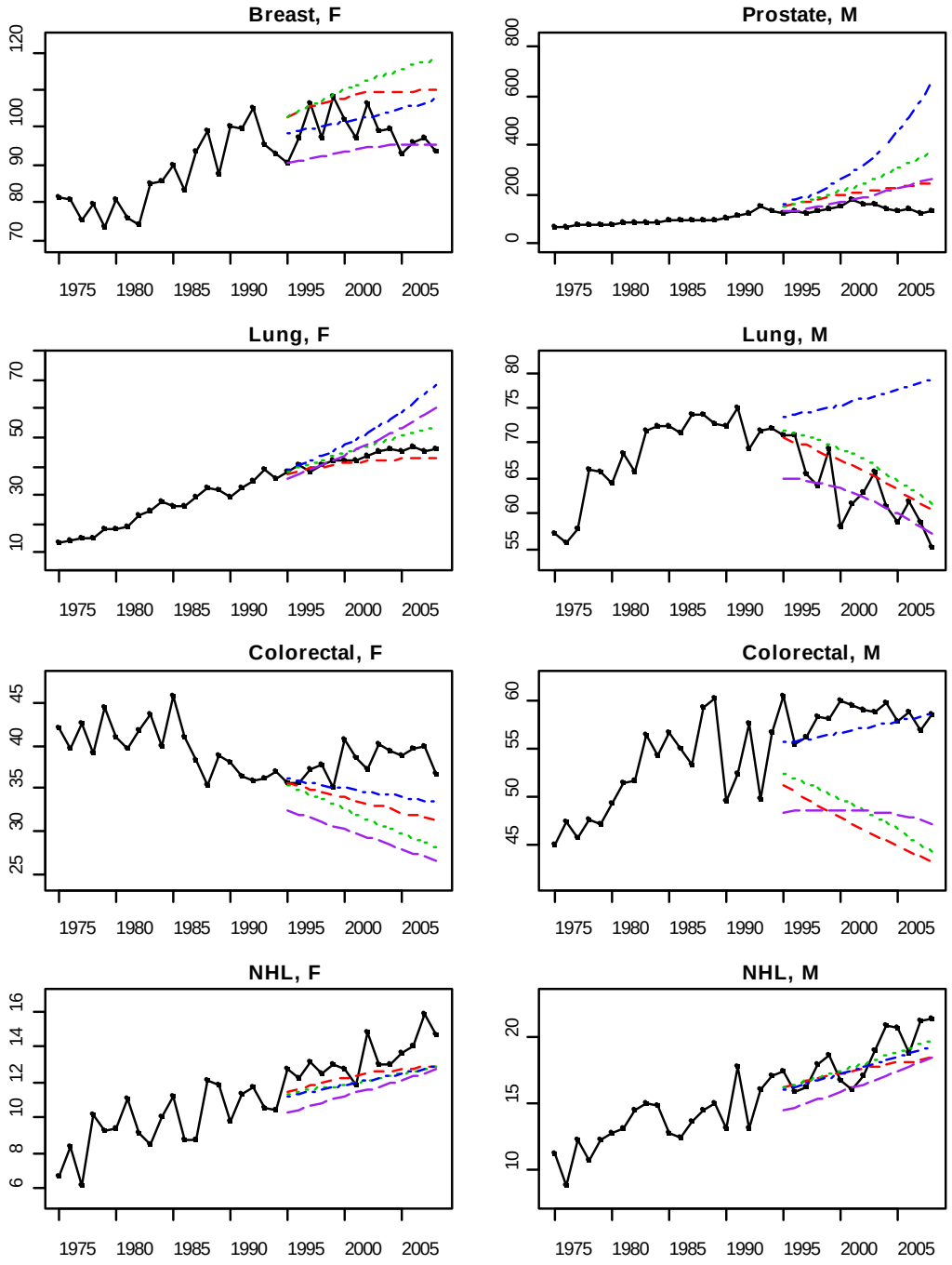

Figure A1. Observed and Projected (1995-2008) ASRs in Alberta, Incidence
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; – • –Blue: HY; – – Purple: NSM

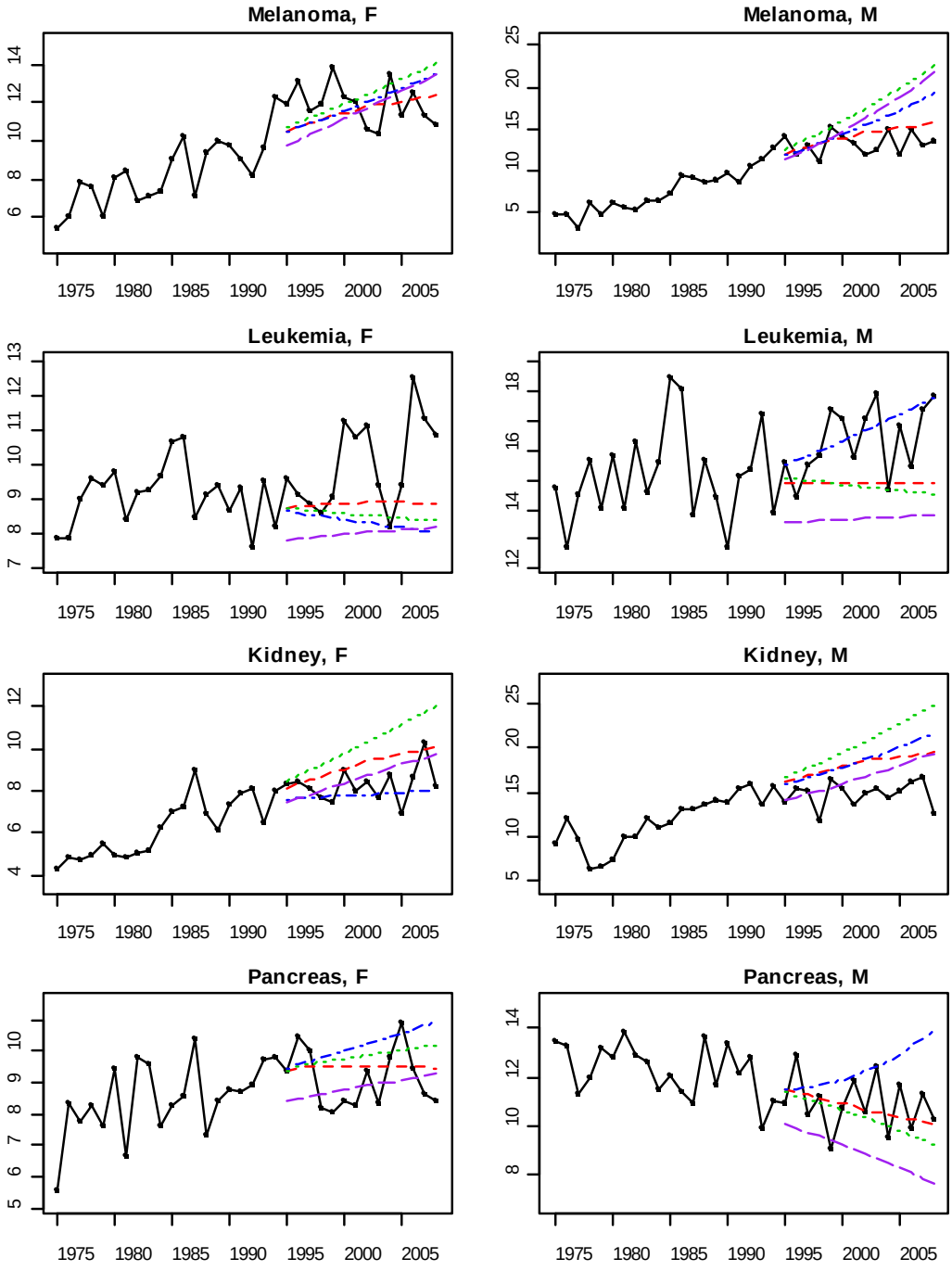

Figure A1 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Incidence
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; –·–Blue: HY; – – Purple: NSM

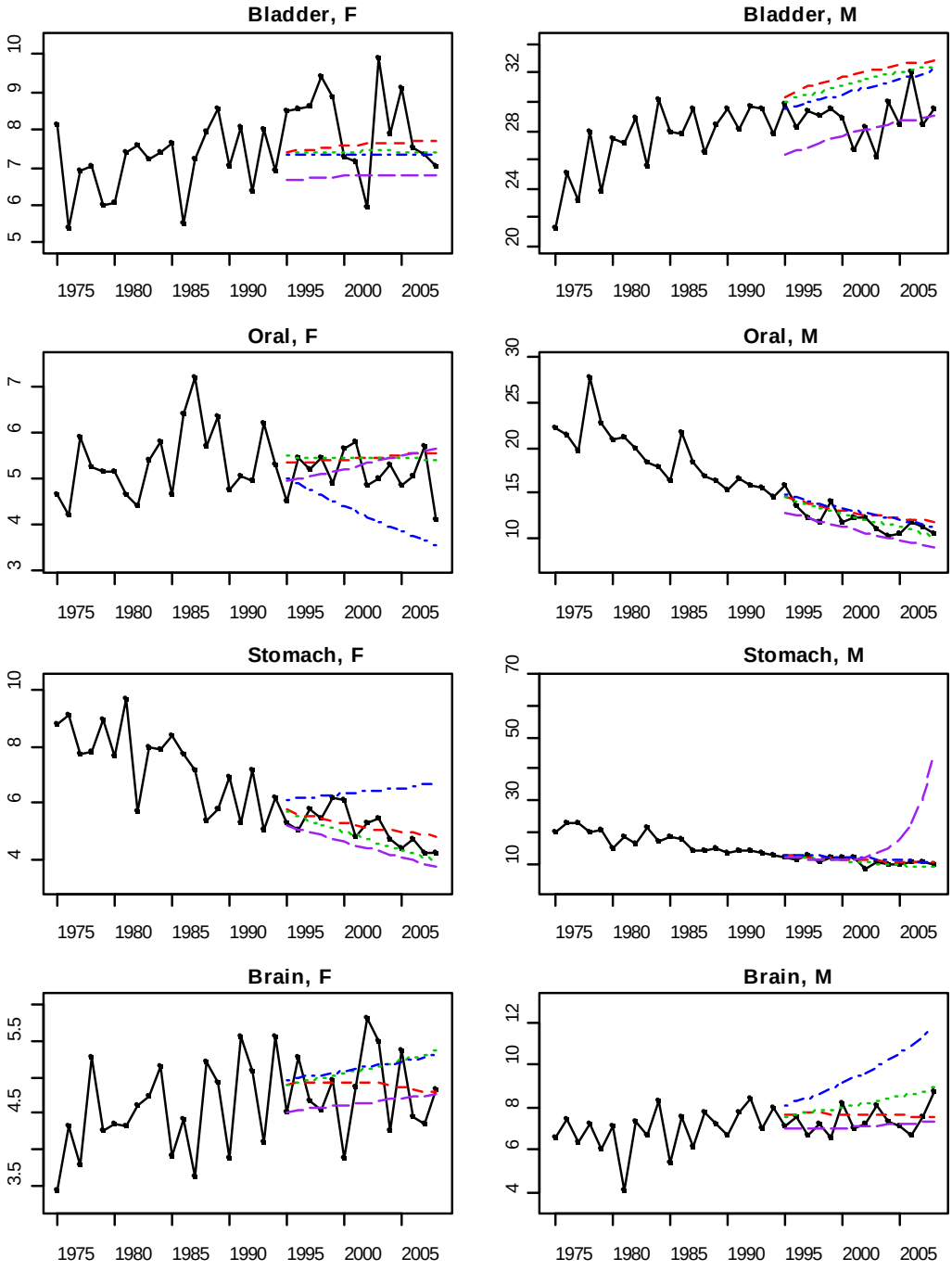

Figure A1 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Incidence
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; – • –Blue: HY; – – Purple: NSM

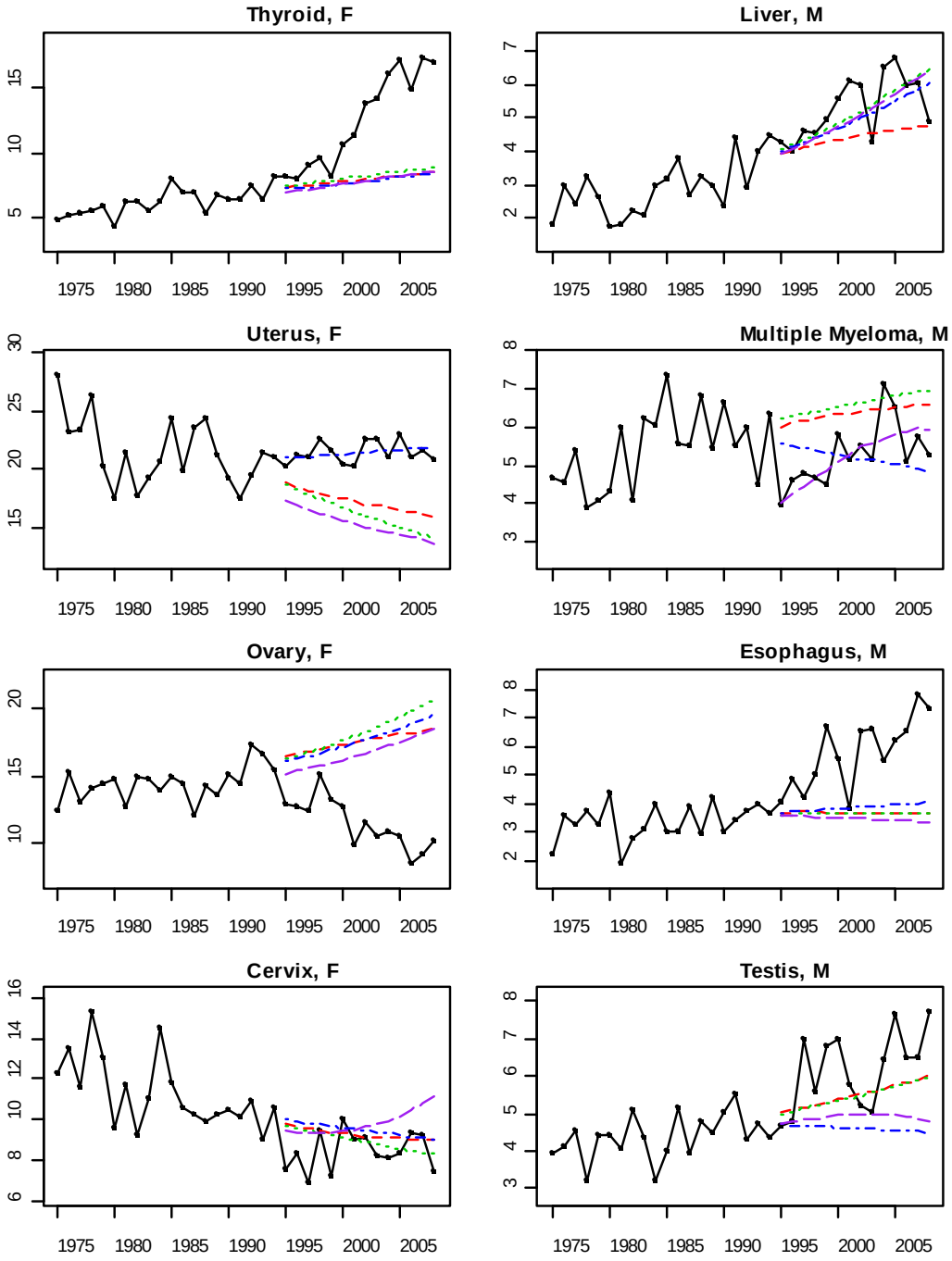

Figure A1 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Incidence
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; · · · Green: BM1; – • –Blue: HY; – – Purple: NSM *
* NSM produced extremely large values and PM projections were presented for male multiple myeloma.

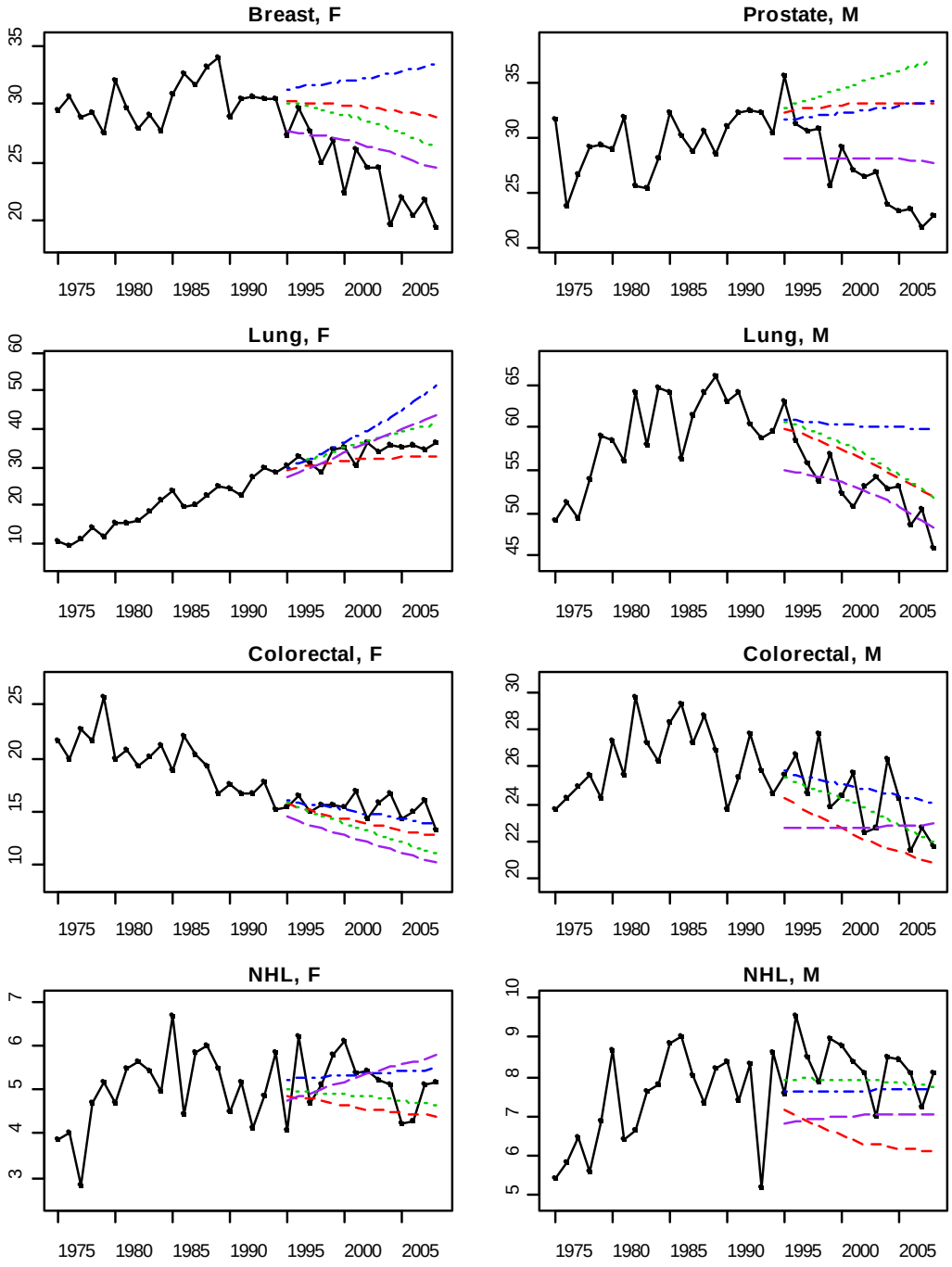

Figure A2. Observed and Projected (1995-2008) ASRs in Alberta, Mortality
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; – • –Blue: HY; – – Purple: NSM

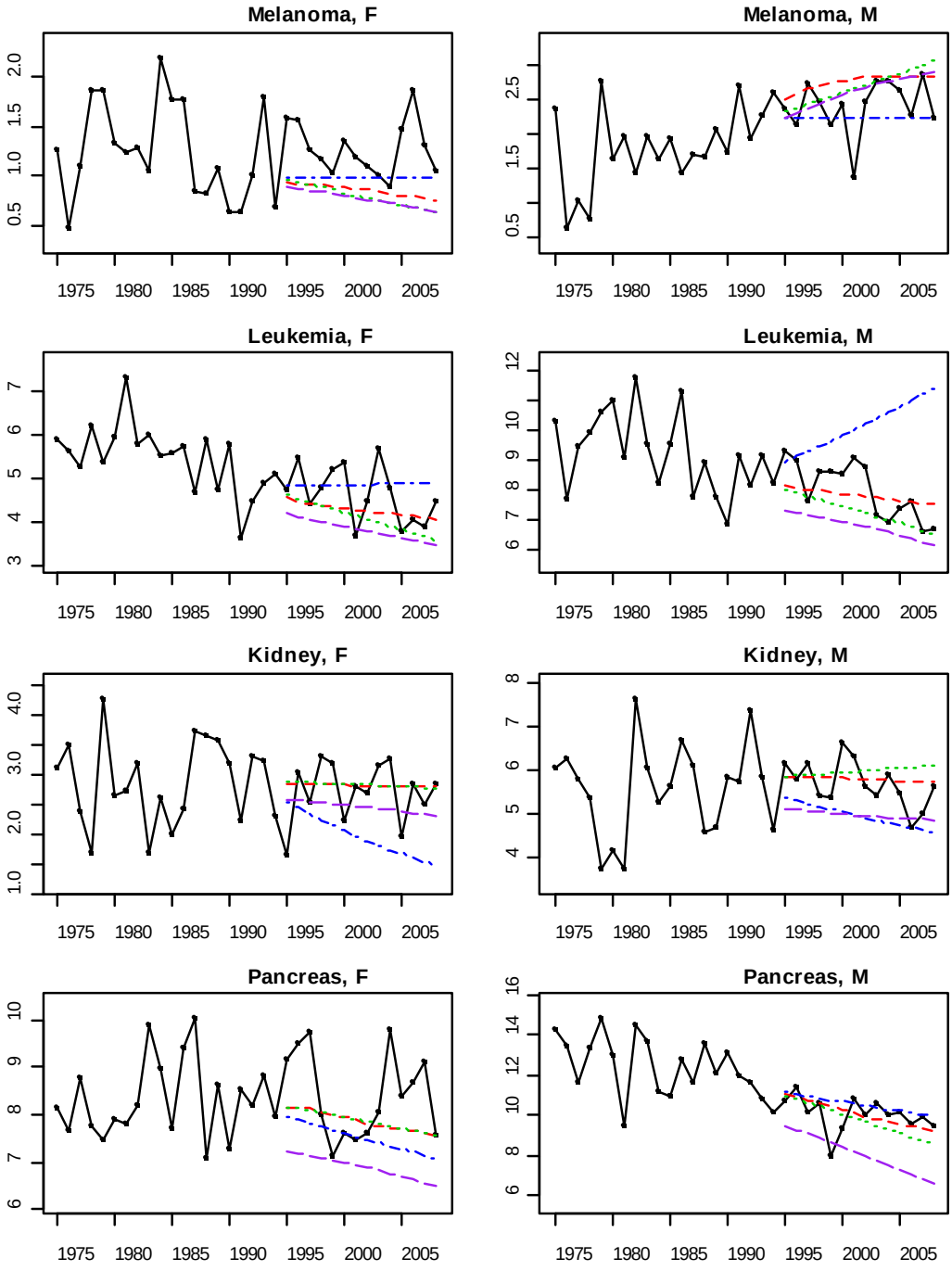

Figure A2 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Mortality
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; –·–Blue: HY; – – Purple: NSM

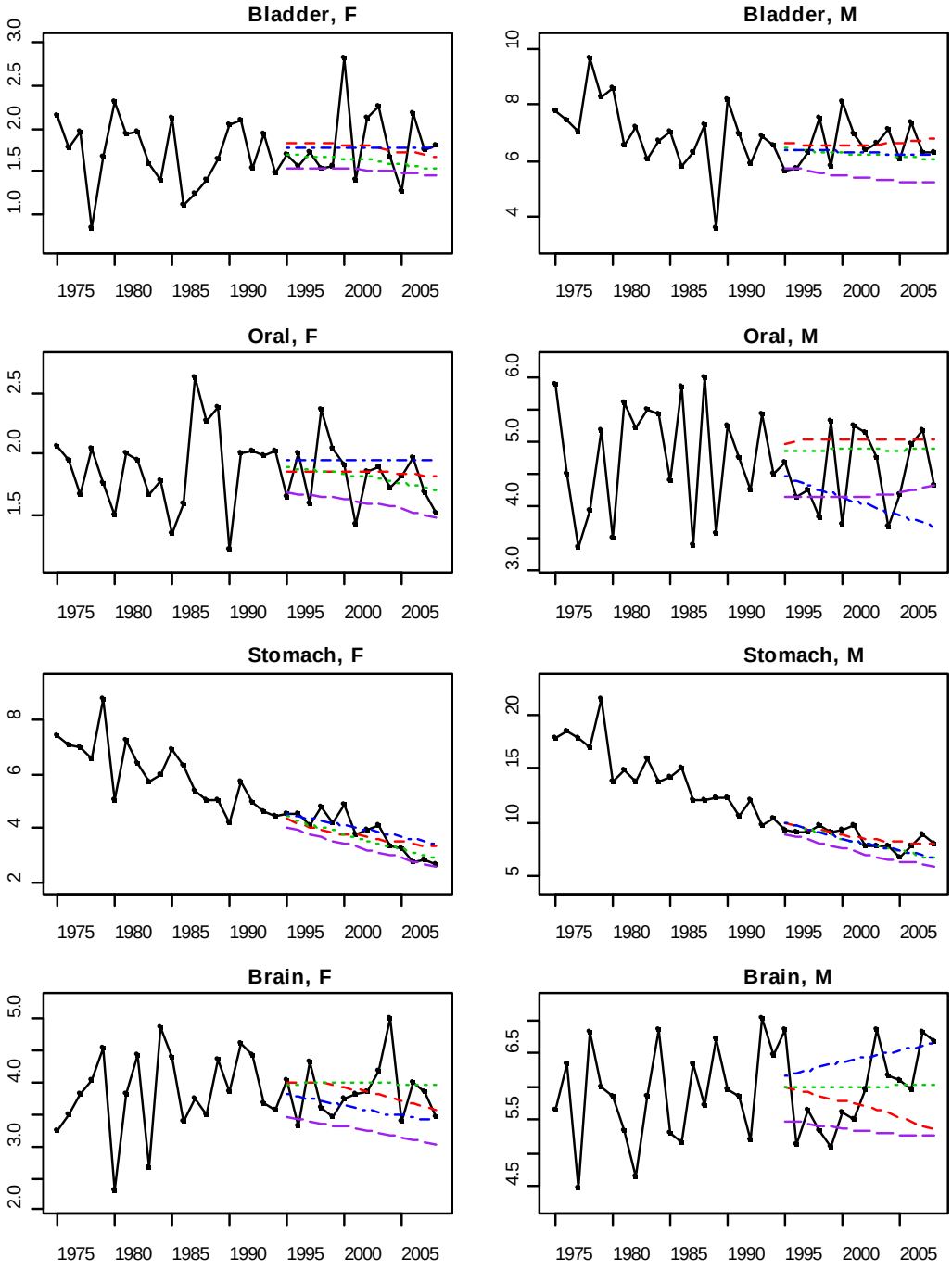

Figure A2 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Mortality
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; ··· Green: BM1; – • –Blue: HY; – – Purple: NSM

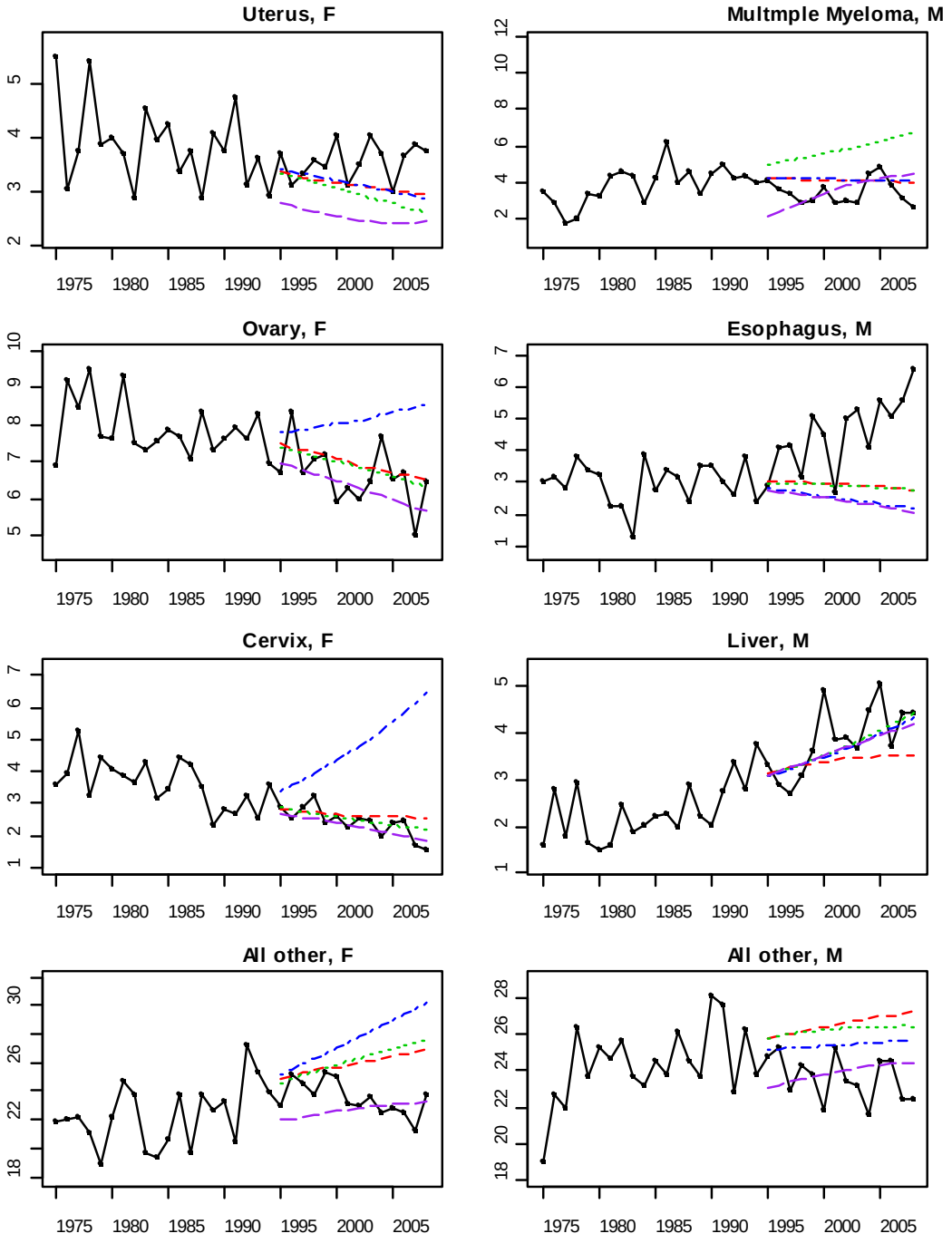

Figure A2 (Continue). Observed and Projected (1995-2008) ASRs in Alberta, Mortality
Left panels: Female; Right panels: Male; — Black: Observed;
– – Red: NP2; · · · Green: BM1; – • –Blue: HY; – – Purple: NSM *

* NSM produced extremely large values and PM projections were presented for male multiple myeloma.

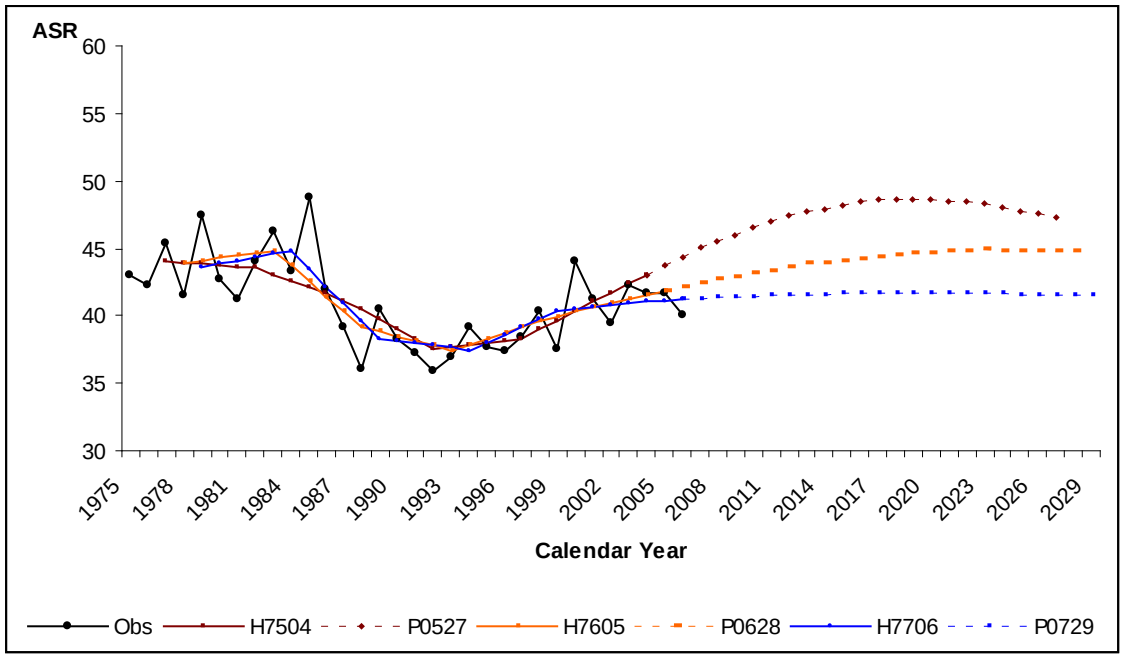

Figure A3. Projections based on different historical data, Colorectal incidence, Female, Alberta
Obs: Observed ASRs; H: Fitted ASRs in historical data 1975-2004, 1976-2005 and 1977-2006;
P: Projected ASRs in 2005-2027, 2006-2028 and 2007-2029

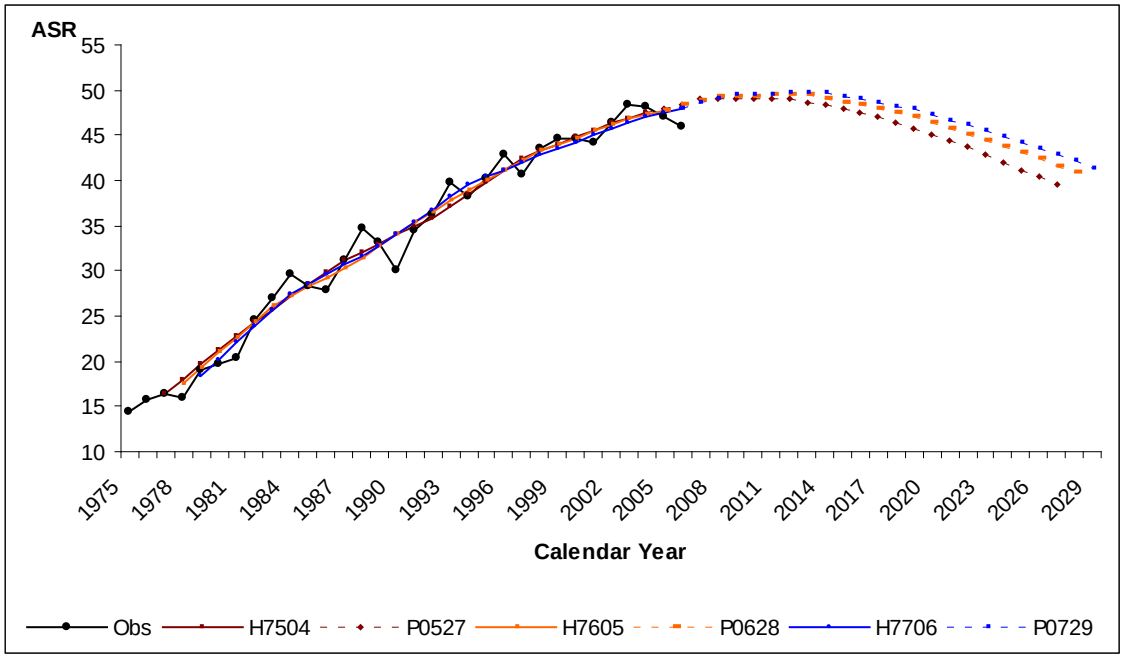

Figure A4. Projections based on different historical data, Lung incidence, Female, Alberta
Obs: Observed ASRs; H: Fitted ASRs in historical data 1975-2004, 1976-2005 and 1977-2006;
P: Projected ASRs in 2005-2027, 2006-2028 and 2007-2029

Table A1: Cramér-von-Mises test for *RB*s in 1995-2008, Incidence, Alberta

Cancer Type	Female			Male		
	HY *vs* NP2	NP2 *vs* BM1	BM1 *vs* HY	HY *vs* NP2	NP2 *vs* BM1	BM1 *vs* HY
Prostrate	n/a	n/a	n/a	NP2	=	=
Lung	NP2	NP2	BM1	NP2	=	BM1
Breast	=	=	HY	n/a	n/a	n/a
Colorectal	=	=	HY	HY	=	HY
NHL	=	=	=	=	=	=
Melanoma	=	=	=	=	NP2	=
Leukemia	=	=	=	=	=	HY
Kidney	=	=	HY	=	NP2	=
Pancreas	=	=	=	=	=	=
Bladder	=	=	=	=	=	=
Oral Cancer	NP2	=	BM1	=	=	=
Stomach	NP2	=	BM1	=	=	=
Brain	=	=	=	NP2	=	BM1
Liver	n/a	n/a	n/a	=	=	=
Multiple myeloma	n/a	n/a	n/a	=	=	HY
Esophagus	n/a	n/a	n/a	=	=	=
Thyroid	=	=	=	n/a	n/a	n/a
Body of Uterus	HY	=	HY	n/a	n/a	n/a
Ovary	=	=	=	n/a	n/a	n/a
Cervix	=	=	=	n/a	n/a	n/a
Testis	n/a	n/a	n/a	NP2	=	BM1
All Other	=	=	=	=	=	=
All Cancers	=	=	=	NP2	=	BM1

=: Statistically insignificant difference in *RB*s between the two projection methods;

Table A2: Cramér-von-Mises test for *RB*s in 1995-2008, Mortality, Alberta

Cancer Type	Female			Male		
	HY *vs* NP2	NP2 *vs* BM1	BM1 *vs* HY	HY *vs* NP2	NP2 *vs* BM1	BM1 *vs* HY
Prostrate	n/a	n/a	n/a	=	=	=
Lung	HY	=	=	NP2	=	BM1
Breast	=	=	=	n/a	n/a	n/a
Colorectal	=	=	HY	=	=	=
NHL	=	=	=	HY	BM1	=
Melanoma	=	=	HY	=	=	=
Leukemia	=	=	=	NP2	=	BM1
Kidney	NP2	=	BM1	NP2	=	=
Pancreas	=	=	=	=	=	=
Bladder	=	=	=	=	=	=
Oral Cancer	=	=	=	=	=	=
Stomach	=	=	=	=	=	=
Brain	=	=	=	=	=	=
Liver	n/a	n/a	n/a	=	=	=
Multiple myeloma	n/a	n/a	n/a	=	NP2	HY
Esophagus	n/a	n/a	n/a	=	=	=
Body of Uterus	=	=	=	n/a	n/a	n/a
Ovary	NP2	=	BM1	n/a	n/a	n/a
Cervix	NP2	=	BM1	n/a	n/a	n/a
All Others	=	=	=	=	=	=
All Cancers	NP2	BM1	BM1	=	BM1	BM1

=: Statistically insignificant difference in *RB*s between the two projection methods;

Table A3: Selected model from age-period-cohort analysis and the Cramér-von-Mises statistics *T* for projected ASRs and observed ASRs in Alberta 1995-2008, Incidence, Female

Cancer Type	Model[1]	HY	NP1	NP2	NP3	PM	NSM	BM1	BM2	BM3
Lung	APC	0.78	***0.29***	***0.35***	***0.47***	***0.44***	***0.38***	***0.44***	0.83	2.34
Breast	AD	*0.73*	1.7	1.61	1.11	1.25	1.19	1.78	1.74	2.34
Colorectal	AD	1.83	2.27	2.21	2.21	2.21	2.34	2.27	2.09	1.3
Non-Hodgkin Lymphoma	AD	1.32	***0.29***	0.91	0.91	2.34	1.69	1.26	***0.25***	2.34
Melanoma	AD	***0.02***	***0.05***	***0.12***	***0.12***	2.34	***0.08***	***0.1***	***0.22***	2.34
Leukemia	A	1.79	1.15	1.15	1.15	2.34	2.34	1.58	1.22	1.48
Kidney	AD	0.86	1.31	1.03	1.03	2.34	***0.19***	1.66	1.79	2.34
Pancreas	AD	1.06	0.88	0.86	0.86	0.89	*0.49*	0.85	1.14	2.21
Bladder (+ in situ)	A	*0.69*	***0.45***	*0.55*	*0.55*	*0.48*	1.87	*0.73*	*0.55*	1.16
Oral	A	1.36	*0.71*	*0.66*	*0.66*	2.34	***0.25***	*0.66*	1.68	*0.5*
Stomach	AD	2.21	***0.14***	***0.23***	***0.23***	***0.09***	*0.7*	***0.17***	***0.07***	2.34
Brain	A	0.89	*0.57*	*0.51*	*0.51*	***0.44***	***0.45***	0.8	0.79	1.01
Thyroid	AD	1.74	1.43	1.5	1.5	2.34	1.57	1.29	1.08	2.34
Body of Uterus	APC	***0.22***	2.34	2.34	2.34	2.34	2.34	2.34	2.34	0.81
Ovary	AD	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34	1.87
Cervix	AD	1.02	0.76	*0.62*	*0.62*	1.68	1.61	***0.35***	***0.12***	2.34
All Other Cancers	AD	***0.34***	***0.18***	***0.25***	***0.25***	1.4	1.88	*0.4*	***0.09***	2.34
All Cancers	AD	***0.34***	***0.04***	***0.2***	***0.2***	2.34	2.1	***0.12***	***0.25***	2.34

[1]. Selected models from age-period-cohort analysis in 1975-1994;
[*] Bold italic font represents $T \leq 0.47$, where the null hypothesis of equal distributions is not rejected at $\alpha = 0.05$;
[**] Italic font represents $T \leq 0.73$.

Table A3 (Continue): Selected model from age-period-cohort analysis and the Cramér-von-Mises statistics *T* for projected ASRs and observed ASRs in Alberta 1995-2008, Incidence, Male

Cancer Type	Model[1]	HY	NP1	NP2	NP3	PM	NSM	BM1	BM2	BM3
Prostate	APC	2.15	2.09	1.99	1.35	2.34	1.08	2.04	2.04	2.34
Lung	AC	2.34	***0.39***	***0.37***	1.57	***0.32***	***0.13***	*0.6*	1.33	***0.24***
Colorectal	AP	0.9	2.34	2.34	2.34	2.34	2.34	2.34	2.34	2.34
Non-Hodgkin Lymphoma	AD	***0.16***	***0.12***	***0.34***	***0.34***	*0.64*	0.79	***0.12***	***0.1***	2.34
Melanoma	AD	*0.6*	0.85	***0.32***	***0.32***	2.34	*0.61*	1.22	1.24	2.34
Leukemia	A	***0.19***	1.48	1.48	1.48	2.34	2.34	1.48	*0.72*	1.95
Kidney	AD	1.98	2.21	2.04	2.04	2.04	*0.64*	2.27	2.34	2.34
Pancreas	A	1.14	***0.16***	***0.12***	***0.12***	***0.37***	1.71	***0.21***	***0.14***	1.58
Bladder (+ in situ)	AD	1.6	2.06	1.95	1.95	2.34	*0.59*	1.83	1.98	1.67
Oral	AD	***0.37***	***0.27***	*0.65*	*0.65*	***0.1***	***0.44***	***0.08***	***0.1***	2.34
Stomach	AD	*0.71*	***0.09***	***0.27***	***0.27***	***0.1***	0.99	***0.07***	***0.07***	2.34
Brain	A	1.99	1.16	1.16	1.16	0.9	***0.25***	1.24	1.17	1.1
Liver	AD	***0.19***	***0.4***	0.78	0.78	2.34	***0.09***	***0.05***	***0.14***	2.34
Multiple Myeloma	AD	***0.2***	1.55	1.48	1.48	*0.64*	1.43	1.58	1.76	*0.66*
Esophagus	A	1.91	2.34	2.34	2.34	n/a[†]	2.34	2.34	2.06	2.34
Testis	AD	2.21	*0.52*	*0.58*	*0.58*	*0.64*	1.48	*0.58*	*0.7*	2.34
All Other Cancers	AD	1.15	*0.55*	*0.69*	*0.69*	*0.66*	1.87	*0.55*	*0.56*	2.27
All Cancers	AD	2.15	1.06	*0.73*	*0.73*	1.07	1.44	1.21	1.15	2.34

[1]. Selected models from age-period-cohort analysis in 1975-1994;
[*] Bold italic font represents $T \leq 0.47$, where the null hypothesis of equal distributions is not rejected at α =0.05;
[**] Italic font represents $T \leq 0.73$;
[†] The average method (only age effects) was used and thus the projected ASRs are the same value.

Table A4: Selected model from age-period-cohort analysis and the Cramér-von-Mises statistics *T* for projected ASRs and observed ASRs in Alberta 1995-2008, Mortality, Female

Cancer Type	Model[1]	HY	NP1	NP2	NP3	PM	NSM	BM1	BM2	BM3
Lung	APC	*0.69*	*0.65*	0.81	***0.23***	*0.58*	***0.3***	***0.42***	0.85	2.34
Breast	A	2.34	2.02	2.02	2.02	1.9	*0.57*	1.43	2.06	2.34
Colorectal	AD	***0.26***	1.09	0.88	0.88	1.06	1.99	1.11	0.9	2.34
Non-Hodgkin Lymphoma	A	*0.65*	1.29	0.99	0.98	2.27	***0.15***	0.89	*0.71*	0.88
Melanoma	A	1.87	2.06	2.06	2.06	2.34	2.34	2.16	2.16	***0.43***
Leukemia	AD	0.91	*0.58*	*0.65*	*0.65*	1.74	1.16	*0.64*	***0.45***	1.65
Kidney	A	1.28	***0.34***	***0.42***	***0.42***	2.34	0.81	***0.35***	*0.73*	*0.73*
Pancreas	A	0.89	***0.45***	***0.42***	***0.42***	***0.42***	2.16	***0.44***	*0.59*	*0.73*
Bladder (+ in situ)	A	*0.73*	***0.42***	***0.38***	***0.38***	***0.24***	1.48	*0.51*	*0.59*	*0.69*
Oral	A	0.91	*0.5*	***0.43***	***0.43***	2.34	0.92	***0.34***	0.91	0.91
Stomach	AD	***0.18***	***0.29***	***0.29***	***0.29***	1.99	*0.58*	***0.21***	***0.18***	2.34
Brain	A	*0.54*	***0.19***	***0.18***	***0.18***	*0.73*	1.94	*0.73*	0.9	*0.73*
Body of Uterus	A	1.16	1.52	1.19	1.19	1.42	2.34	1.45	1.1	1.38
Ovary	A	1.89	***0.28***	*0.5*	*0.5*	0.91	***0.26***	***0.25***	0.78	1.87
Cervix	AD	2.34	0.91	0.91	0.91	0.94	***0.21***	***0.07***	***0.14***	2.34
All Other Cancers	AD	2.21	2.04	1.98	1.98	1.88	*0.56*	1.83	1.6	1.76
All Cancers	AD	2.1	1.55	1.48	1.48	0.87	2.34	***0.1***	***0.41***	0.9

[1]. Selected models from age-period-cohort analysis in 1975-1994;
[*] Bold italic font represents $T \leq 0.47$, where the null hypothesis of equal distributions is not rejected at $\alpha = 0.05$;
[**] Italic font represents $T \leq 0.73$.

Table A4 (Continue): Selected model from age-period-cohort analysis and the Cramér-von-Mises statistics *T* for projected ASRs and observed ASRs in Alberta 1995-2008, Mortality, Male

Cancer Type	Model[1]	HY	NP1	NP2	NP3	PM	NSM	BM1	BM2	BM3
Prostate	AD	1.87	1.87	1.87	1.87	1.87	*0.73*	1.93	1.91	*0.7*
Lung	APC	1.87	***0.33***	*0.51*	1.76	1.49	***0.08***	*0.66*	1.46	***0.39***
Colorectal	A	***0.34***	1.2	0.82	1.16	1.94	*0.73*	***0.16***	0.86	1.15
Non-Hodgkin Lymphoma	AD	1.16	2.21	2.21	***0.35***	2.34	1.91	0.89	***0.37***	1.93
Melanoma	A	1.16	2.16	1.13	1.13	***0.31***	***0.28***	*0.49*	*0.49*	1.87
Leukemia	AD	2.09	***0.37***	***0.47***	***0.47***	2.34	1.21	***0.44***	***0.37***	2.27
Kidney	A	1.53	*0.68*	*0.6*	*0.6*	*0.73*	1.66	0.82	0.79	0.96
Pancreas	AD	***0.37***	***0.08***	***0.03***	***0.03***	***0.34***	1.86	***0.12***	***0.04***	2.34
Bladder (+ in situ)	AD	***0.28***	*0.54*	*0.55*	*0.55*	2.27	2.15	***0.31***	0.99	1.39
Oral	A	*0.53*	0.86	0.86	0.86	*0.73*	*0.6*	*0.73*	*0.71*	*0.56*
Stomach	AD	***0.16***	***0.11***	***0.18***	***0.18***	***0.24***	0.94	***0.17***	***0.21***	2.34
Brain	A	0.91	***0.35***	***0.4***	***0.4***	*0.62*	1.22	*0.62*	*0.62*	*0.62*
Liver	AD	***0.17***	*0.59*	0.88	0.88	2.34	***0.18***	***0.12***	***0.19***	2.34
Esophagus	A	2.16	1.62	1.51	1.51	1.15	2.27	1.53	1.48	1.48
Multiple Myeloma	AD	1.38	1.48	1.16	2.27	2.1	2.15	2.34	2.34	***0.35***
All Other Cancers	AD	1.99	2.34	2.34	2.34	2.34	***0.3***	2.34	2.34	*0.51*
All Cancers	A	1.51	1.87	1.87	1.87	***0.12***	1.35	0.84	1.45	1.87

[1]. Selected models from age-period-cohort analysis in 1975-1994;

[*] Bold italic font represents $T \leq 0.47$, where the null hypothesis of equal distributions is not rejected at $\alpha = 0.05$;

[**] Italic font represents $T \leq 0.73$.

Appendix B. Details of Bayesian Methods

B.1. Priors of Age, Period and Cohort Effects:

Suppose that we plan to get the projections for *M* future periods and cohorts, the evolving prior distributions for three factorial parameters can be defined as follows (Bray, 2002):

BM1.

i. For the *A* age effects:

$$\alpha_1 \sim N(0,10^6 \times \sigma_\alpha^2),$$

$$\alpha_2 \mid \alpha_1 \sim N(0,10^6 \times \sigma_\alpha^2),$$

$$\alpha_a \mid \alpha_1,\cdots,\alpha_{a-1} \sim N(2\alpha_{a-1} - \alpha_{a-2}, \sigma_\alpha^2), \quad 3 \le a \le A;$$

ii. For the *P* + *M* period effects:

$$\beta_1 \sim N(0,10^6 \times \sigma_\beta^2),$$

$$\beta_2 \mid \beta_1 \sim N(0,10^6 \times \sigma_\beta^2),$$

$$\beta_p \mid \beta_1,\cdots,\beta_{p-1} \sim N(2\beta_{p-1} - \beta_{p-2}, \sigma_\beta^2), \quad 3 \le p \le P + M;$$

iii. For the *C* + *M* cohort effects:

$$\gamma_1 \sim N(0,10^6 \times \sigma_\gamma^2),$$

$$\gamma_2 \mid \gamma_1 \sim N(0,10^6 \times \sigma_\gamma^2),$$

$$\gamma_c \mid \gamma_1,\cdots,\gamma_{\gamma-1} \sim N(2\gamma_{c-1} - \gamma_{c-2}, \sigma_\gamma^2), \quad 3 \le c \le C + M.$$

The Bayesian age-period modeling methods in short-term projection (Qiu, Jiang & Hatcher, 2010) can be extended to Bayesian APC modeling method. And the priors of the model are specified as follows:

BM2.

i. For the *A* age effects:

$$\alpha_1 = 0, \alpha_a \sim N(0,10^6), \qquad 2 \le a \le A;$$

ii. For the *P* + *M* period effects: $\beta_p \sim N(\overline{\beta}_p, \sigma_p^2), p = 1,2,\cdots,P+M,$ where

$$\overline{\beta}_1 = (2\beta_2 - \beta_3)/K_1, K_1 = 1,$$

$\overline{\beta}_2 = (2\beta_1 + 4\beta_3 - \beta_4)/K_2, K_2 = 5,$

$\overline{\beta}_p = (4\beta_{p-1} + 4\beta_{p+1} - \beta_{p-2} - \beta_{p+2})/K_p, K_p = 6, \ \ 3 \le p \le P + M - 2,$

$\overline{\beta}_{P+M-1} = (2\beta_{P+M} + 4\beta_{P+M-2} - \beta_{P+M-3})/K_{P+M-1}, K_{P+M-1} = 5,$

$\overline{\beta}_{P+M} = (2\beta_{P+M-1} - \beta_{P+M-2})/K_{P+M}, K_{P+M} = 1,$ and

$\sigma_p^2 = u^2/K_p, p = 1,2,\cdots,P+M$;

iii. For the $C + M$ cohort effects: $\gamma_c \sim N(\overline{\gamma}_c, \sigma_c^2), c = 1,2,\cdots,C+M,$ where

$\overline{\gamma}_1 = (2\gamma_2 - \gamma_3)/K_1, K_1 = 1,$

$\overline{\gamma}_2 = (2\gamma_1 + 4\gamma_3 - \gamma_4)/K_2, K_2 = 5,$

$\overline{\gamma}_c = (4\gamma_{c-1} + 4\gamma_{c+1} - \gamma_{c-2} - \gamma_{c+2})/K_c, K_c = 6, \ \ 3 \le c \le C + M - 2,$

$\overline{\gamma}_{C+M-1} = (2\gamma_{C+M} + 4\gamma_{C+M-2} - \gamma_{C+M-3})/K_{C+M-1}, K_{C+M-1} = 5,$

$\overline{\gamma}_{C+M} = (2\gamma_{C+M-1} - \gamma_{C+M-2})/K_{C+M}, K_{C+M} = 1,$ and

$\sigma_c^2 = v^2/K_c, c = 1,2,\cdots,C+M$.

The model is an extension of the model of Breslow & Clayton (1993) used to analyze the "Iceland Breast Cancer Data" under Bayesian framework (see "Examples Volume II" under the WinBUGS "Help" tab). Because only two hyper-parameter (σ^2) needs to be specified, the computation time is faster than BM1. Clèries, Ribes & Esteban *et al* (2006) proposed the alternative approach, where the prior of the model is specified as follows:

BM3.

i. For the A age effects: $\alpha_a \sim N(\overline{\alpha}_a, \sigma_\alpha^2), a = 1,2,\cdots,A,$ where

$\overline{\alpha}_1 = 2\alpha_2 - \alpha_3,$

$\overline{\alpha}_2 = (2\alpha_1 + 4\alpha_3 - \alpha_4)/5,$

$\overline{\alpha}_a = (4\alpha_{a-1} + 4\alpha_{a+1} - \alpha_{a-2} - \alpha_{a+2})/6, \ \ 3 \le a \le A - 2,$

$\overline{\alpha}_{A-1} = (2\alpha_A + 4\alpha_{A-2} - \alpha_{A-3})/5,$

$\overline{\alpha}_A = 2\alpha_{A-1} - \alpha_{A-2};$

ii. For the $P + M$ period effects (Corner constraint on the first period):

$\beta_1 = 0, \beta_p \sim N(0, \sigma_\beta^2), \qquad 2 \le p \le P + M;$

iii. For the $C + M$ cohort effects (Corner constraint on the first cohort):

$$\gamma_1 = 0, \gamma_c \sim N(0, \sigma_\gamma^2), \qquad 2 \le c \le C + M .$$

B.2. Hyper-Priors of Variance Parameters:

Different prior distributions for hyper-parameter can be used in Bayesian analysis (e.g. Gelman, 2006). We adopt the conjugate priors for variance parameters: σ_α^2, σ_β^2 and σ_γ^2, for BM1:

$$\sigma_\alpha^2 \sim IG(10^{-4}, 10^{-4}), \quad \sigma_\beta^2 \sim IG(10^{-4}, 10^{-4}) \text{ and } \quad \sigma_\gamma^2 \sim IG(10^{-4}, 10^{-4})$$

where *IG* represents the inverse Gamma distribution. For BM2, we use vague informative conjugate priors for variance parameters: u^2 and v^2:

$$\tau_1 = 1/u^2 \sim Gamma(r_1, d_1) \text{ and } \tau_2 = 1/v^2 \sim Gamma(r_2, d_2)$$

where $r_1 = 0.0001 + (P + N)/2$ and $d_1 = 0.0001 + \sum_{p=0}^{P+N} L_p / 2$ with $L_p = K_p \times \beta_p \times (\beta_p - \bar{\beta}_p)$; and $r_2 = 0.0001 + (C + N)/2$ and $d_2 = 0.0001 + \sum_{c=0}^{C+N} L_c / 2$ with $L_c = K_c \times \gamma_c \times (\gamma_c - \bar{\gamma}_c)$, noting that the precision parameters, τ_1 and τ_2, instead of the variance parameter, u^2 and v^2, are used for normal distribution in WinBUGS. For BM3, we use the approximated Bayesian priors (i.e. uniform prior distribution) for variance components σ_α, σ_β and σ_γ:

$$\sigma_\alpha \sim U(0.1, 10^2), \quad \sigma_\beta \sim U(0.1, 10^2) \text{ and } \quad \sigma_\gamma \sim U(0.1, 10^2) .$$

The conjugate priors for BM1 are chosen such that the type of the posterior distribution of the variance parameter is well-defined. The informative priors for BM2 will accelerate the MCMC running assuming the prior information is correct. The uniform priors in BM3 require more burning-in running as it is hard to track the distribution shape for the posteriors. In WinBUGS, the adapting method is used for the first 500 simulation loops to look for the trace of the distribution shape and those 500 simulated values can not be utilized for inference.

It should be noted that the period effects are extended into projection periods by the "random walk" transitive prior setting in BM2, and the linear trend from the age effects should be constrained in BM1 and BM2. In general, BM1 work for any situation, but BM2 and BM3 may be affected by the number of cases in each age group. For example, when there is no cases in age 0-9 (age group 1 and 2 in the 18 age group setting) in any historical year, we should assume that $\alpha_1 = 0, \alpha_2 = 0$ and

set the initial values as $\alpha_1 = NA, \alpha_2 = NA$, where *NA* represents the missing value in WinBUGS, which is the same as the coding in R/Splus. That means that we essentially assume that it is impossible to observe the cancer cases in those age groups. We will conduct the sensitivity analysis (MacNab, Qiu & Gustafson *et al*, 2004; Qiu, Song & Tan, 2002) to compare the three methods.

B.3. Bayesian Inference and Markov Chain Monte Carlo

When the likelihood function is formed for observed data *D*:

$$P(D \mid \theta)P(\theta),$$

given the prior distribution for parameters θ in the model: $P(\theta)$, the posterior distribution can be solved using Bayes Theorem:

$$P(\theta \mid D) = \frac{P(\theta)P(D \mid \theta)}{\int P(\theta)P(D \mid \theta)d\theta}.$$

The posterior distribution is the objective of all Bayesian inference. Then any features of the posterior distribution, such as moments, quantiles and modes, can be expressed in term of posterior expectations:

$$E[f(\theta) \mid D] = \frac{\int f(\theta)P(\theta)P(D \mid \theta)d\theta}{\int P(\theta)P(D \mid \theta)d\theta}. \qquad \text{(B1)}$$

Only in a few situations the analytic solution for the integral is feasible, e.g. binomial proportion parameter with Beta distribution as its conjugate prior. In most application, analytical evaluation is impossible. When many of the parameters need be estimated, the dimension of the integration can be very high and thus the numerical integration approaches, e.g. Newton-Raphson, Gauss-Hermite quadrature methods, and approximation methods such as Riemann, Laplace and saddlepoint methods, may not work properly. Alternatively, Monte Carlo integration method can be used by drawing large number of independent samples from the distribution function. In details, re-write (B1) as

$$E[f(x)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx}$$

where $\pi(x)$ is a posterior distribution function and x is a parameter or missing data. Then Monte Carlo integration evaluates $E[f(x)]$ by drawing samples (X_t, t = 1, 2, …, n) from $\pi(\circ)$ and then approximation:

$$E[f(x)] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t).$$

When the samples $\{X_t\}$ are independent, the Laws of Large Numbers ensure that the approximation can be made as accurate as desired by increasing the sample size n. In general, drawing samples independently from a distribution is not feasible. In practice, we can generate a sequence of random variable $\{X_0, X_1, X_2, \ldots\}$ such that, given current state X_t, the next state X_{t+1} does not depend on the history of the chain $\{X_0, X_1, \ldots, X_{t-1}\}$. The sequence is called Markov chain. The chain will gradually "forget" the initial state and eventually converge to a unique stationary distribution. This stationarity has been theoretically proved achievable when the Markov chain is geometrically ergodic (Geman & Geman, 1984). After discarding sufficiently long iterations $\{X_t; t = 0, 2, \ldots, m\}$ (called burn-in),

$$E[f(x)] \approx \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t),$$

the points $\{X_t; t = m+1, m+2, \ldots, n\}$ will be dependent samples approximately from $\pi(\circ)$ when Metropolis-Hastings sampling approach is used to construct the Markov chain. A special case of Metropolis-Hastings sampling approach is the Gibbs sampler, which is more efficient with cost of slow convergence and adopted in WinBUGS software.

Unlike numeric convergence of optimal solution by iterative algorithm, stochastic convergence diagnostics should be done to confirm the stationary distribution of samples for each parameter before presenting projection results. That is, before using the chain for Bayesian inference, we should determine whether: (1) the chain converge to a stationary distribution, i.e. m = ? (2) and at what iteration the chain should be stopped, i.e. n = ? WinBUGS provides the BGR diagnostics when multiple chains are run, and some visible diagnostics such as trace, history, quantiles and auto-correlation. The R-package of Convergence Diagnostics and Output Analysis (CODA) is available to implement convergence diagnostics such as the Raftery and Lewis convergence

diagnostic methods, Heidelberger and Welch's stationary test, and Geweke's *Z*-score test: These are:

i. *Convergence to a stationary distribution* (Heidelberger and Welch, 1983): In this method the Kolmogorov-Smirnov statistic based on Brownian Bridge Theory is used to test if sample values for each variable form a stationary process.

ii. *Convergence to i.i.d. sampling* (Raftery and Lewis, 1992): In this method the batch size (*k*, the thinning interval, which only record the k^{th} sample value in each interval of the chain such that the recorded chain is less auto-correlated), the burn-in (*m*), the total run (*n*) and so on by testing if the "thinned" chain is a Markov chain against the 2^{nd} –order Markov chain, are estimated to determine if the independent requirement of simulated values for Monte Carlo integration is met.

iii. *Convergence of average* (Geweke, 1992): To determine if *n* is large enough to exhibit all facets of $\pi(\circ)$, for the first *A*% chain and the last *B*% chain, Geweke defined a *Z*-statistic and suggested $A = 10$, $B = 50$:

$$Z = \frac{\sqrt{n}(\mu_A - \mu_B)}{\sqrt{\sigma_A^2 / n_A + \sigma_B^2 / n_B}} \sim N(0,1).$$

iv. *Convergence of average* (Brooks & Gelman, 1998; Gelman & Rubin, 1992): The Geweke's diagnostic is for single chain. The defect is "you have only seen where you have been". Gelman, Rubin and Brooks suggested two or more parallel chains; each starts from different initial values, which are over-dispersed with respect to the true posterior distribution. Comparing within- and between-chains variances, the "potential scale reduction factor" (or shrink factor), which is *F*-distributed, can be estimated to diagnose the convergence or chains-mixing speed.

Over-dispersed initial values can be issue as the target distribution is not known. In general, one set of initial values for single chain can be the means of prior distributions for every parameter. To create other sets of initial values, the GLM APC model for historical data is first fitted; the estimated age, period and cohort effects with small variance parameters (high precision, say, 10) can be one of the sets of initial values, the opposite values (say, different sign) of estimated effects with large variance parameter (say, precision is 0.01) can be the other set. Usually, the first half of

the iterations should be discarded for each chain. It can be argued that a single chain with $M \times N$ sampling but slow rate of mixing is more likely to get closer to the stationary distribution than M chains of size N. In our practice, the single chain is acceptable to fit the Bayesian APC models if there is no over-parameterization problem in the model (For example, when there is no significant cohort effects for certain data, the cohort effects in the Bayesian APC models are over-parameterized, in such situation, a Bayesian age-period model is recommended). As a pilot study, 4000 preliminary samples can first be generated with $k = 1$ from WinBUGS and use Raftery and Lewis' Diagnostics in R to obtain the suggested thinning intervals (k), burn-in (m) and total run required (n) for target parameters (e.g., ASRs and total counts). Then the sampling in WinBUGS can continue by setting new k and n. After eliminating the burn-in iterations, the remaining draws can be judged to constitute a representative sample for the joint posterior distribution of the model parameters; these are used for posterior inference of the APC models.

Appendix C. Kolmogorov-Smirnov Test and Cramér-von-Mises Test

The Kolmogorov-Smirnov test (KS test) can be used to determine if two datasets differ significantly. It is a non-parametric testing method without distribution assumption. Other tests (e.g. *t*-test) may be more sensitive under the student *t* distribution. The KS-test is a robust test that cares only about the relative distribution of the data. To test whether two underlying one-dimensional probability distributions differ, the Kolmogorov–Smirnov statistic is

$$D = \sup_x | F_{1,m}(x) - F_{2,n}(x) | \qquad \text{(C1)}$$

where $F_{1,\,m}$ and $F_{2,\,n}$ are the empirical distribution functions of the first (with m samples) and the second (with n samples) sample respectively. The null hypothesis is rejected at level α (say, 0.05) if

$$\sqrt{\frac{m \cdot n}{m+n}} D > K_\alpha ,$$

where K_α is the critical value of the Kolmogorov distribution (Kolmogorov, 1933) at the probability $1 - \alpha$, and Smirnov (1948) gave the table of the distribution. Under null hypothesis that the two sample sets come from the same distribution, Kolmogorov gave the asymptotic distribution which is the supremum of the Brownian bridge for large sample size. That is,

$$\sqrt{\frac{m \cdot n}{m+n}} D \xrightarrow{m\to\infty, n\to\infty} \sup_x | B[F_{1,m}(x) - F_{2,n}(x)] | ,$$

where $B(\bullet)$ is the Brownian bridge. Therefore, the KS-test needs relevant large number of samples.

The alternative of the KS test is the Cramér-von-Mises test (CVM test), which may be used for small sample size. Unlike the maximum of distance between two distributions in KS test, the Cramér-von-Mises criterion is defined as the Riemann integration of the squared distance between two distribution functions:

$$w^2 = \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \, dF^*(x) . \qquad \text{(C2)}$$

In one-sample applications F^* is a theoretical distribution and F_n is an empirically observed distribution. In two-sample case, the two distributions can both be empirically estimated distributions (i.e. $F^*(x) = F_m(x)$). Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_m$ be the observed values in the first and second sample respectively, in increasing order; Let $r_1, r_2, \cdots, r_n$ be the ranks of the x's in the combined sample, and let $s_1, s_2, \cdots, s_m$ be the ranks of the y's in the combined sample. Anderson (1962) shows that

$$T = nw^2 = \frac{U}{n \cdot m \cdot (m+n)} - \frac{4m \cdot n - 1}{6(m+n)}, \qquad \text{(C3)}$$

where U is defined as

$$U = n\sum_{i=1}^{n}(r_i - i)^2 + m\sum_{j=1}^{m}(s_j - j)^2. \qquad \text{(C4)}$$

If the value of T is larger than the tabulated value (Anderson, 1962), we can reject the hypothesis that the two samples come from the same distribution. Xiao, Gordon & Yakovlev (2007) implemented the two-sample CVM test and also provided the R-package "CvM2SL2Test" for application. The critical value of CVM statistic T is 0.73, 0.47 and 0.36 for $m = n = 14$ at $\alpha = 0.01$, 0.05 and 0.1, respectively. For large sample size m and n, the computation is slow. When there are many ties between the two samples, e.g. two exact same samples or one sample with the same values (e.g. projected rates by average method), the KS test can not produce correct p-value while the CVM test can not process the computation.

References

Anderson, TW. On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, 1962, **33** (3): 1148–1159.

Baker A & Bray I. Bayesian projections: What are the effects of excluding data from younger age groups? *American Journal of Epidemiology*, 2005, **162**: 798 - 805.

Boyle P, d'Onofrio A & Maisonneuve P *et al*. Measuring progress against cancer in Europe : has the 15% decline targeted for 2000 come about ? *Annals of Oncology*, 2003, **14**: 1312 – 25.

Bray F & Møller B. Predicting the future burden of cancer. *Nature Reviews*, 2006, **6**: 63 - 74.

Bray I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *J R Stat Soc Ser C, Appl Stat*, 2002, **51**: 151 - 64.

Breslow NE and Day NE. *Statistical Methods in Cancer Research*. London: Oxford University Press, 1987.

Breslow NE & Clayton DG. Approximate inference in generalized linear mixed models. *J Am Stat Ass*, 1993, **88**(421): 9 - 25.

Brooks S & Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Statist*, 1998, **7**(4): 434 – 455.

Canadian Cancer Society, Statistics Canada, Provincial / Territorial Cancer Registries & Public Health Agency of Canada. *Canadian Cancer Statistics, 2009* - available at www.cancer.ca.

Carstensen B. Age-period-cohort models for the lexis diagram. *Stat Med*, 2007, **26**(15): 3018 - 45.

Case RM. Cohort analysis of mortality rates as an historical or narrative technique. *British Journal of Preventative and Social Medicine*, 1956, **10**: 159 - 71.

Clayton D & Schifflers E. Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Stat Med*, 1987a, **6**: 449 - 467.

Clayton D & Schifflers E. Models for temporal variation in cancer rates. II. Age-period-cohort models. *Stat Med*, 1987b, **6**: 469-481.

Clements MS, Armstrong BK & Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics*, 2005, **6**(4):576 – 89.

Clèries R, Ribes J & Esteban L *et al*. Time trends of breast cancer mortality in Spain during the period 1977–2001 and Bayesian approach for projections during 2002–2016. *Annals of Oncology*, 2006, **17**: 1783 – 91.

Coleman MP, Esteve J & Damiecki P *et al*. Trends in cancer incidence and mortality. *IARC Scientific Publications*, 1993, **121**: 1 – 806.

De Angelis G, De Angelis R & Frova L *et al*. a computer package to estimate chronic disease morbidity using mortality and survival data. *Compute Methods and Programs in Biomedicine*}, 1994, **44**: 99 - 107.

Decarli A, La Vecchia C & Frova L *et al*. Age, period and cohort models: review of knowledge and implementation in GLIM. *Rev Stat Appl*, 1987, **20**: 397-409.

Dyba T & Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. *Stat Med*, 2000, **19**: 1741 - 52.

Dyba T, Hakulinen T & Päivärinta L. A simple non-linear model in incidence prediction. *Stat Med*, 1997, **16**: 2297 - 309.

Elkum NB. Predicting confidence intervals for the age-period-cohort model. *Journal of Data Science*, 2005, **3**: 403 - 14.

Erbas B, Hyndman RJ & Gertig DM. Forecasting age-specific breast cancer mortality using functional data models. *Stat Med*, 2007, **26**: 458 – 70.

Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 2006, **1**(3): 515 - 33.

Gelman A & Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science*, 1992, **7**: 457 – 511.

Geman S & Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*, 1984, **6**: 721 – 741.

Geweke J. Evaluation the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In Bernardo JM, Berger JO, Dawid AP & Smith AFM Eds. *Bayesian Statistics* 4. Oxford: Oxford University Press, 1992, pp169 – 193.

Hastie TJ & Tibshirani RJ. *Generalized Additive Models*. London: Chapman and Hall, 1990.

Heidelberger P & Welch P. Simulation run length control in the presence of an initial transient. *Operations Research*, 1983, **31**: 1109 – 1144.

Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 1983, **39**: 311 - 324.

Kim HJ, Fay MP & Feuer EJ *et al*. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*, 2000, **19**: 335 - 51 (correction: 2001, **20**:655).

Kolmogorov, A. Sulla determinazione empirica di una legge di distribuzione, *G. Inst. Ital. Attuari*, 1933, **4**: 83.

Lee TCK, Dean CB & Semenciw R. Short-term cancer mortality projections: a comparative study of prediction methods. Technique Report. Simon Fraser University & Public Health Agency of Canada. 2010.

Lunn DJ, Thomas A, Best N & Spiegelhalter D. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 2000, **10**: 325 - 337.

MacNab Y, Qiu Z & Gustafson P *et al*. Hierarchical Bayes analysis of multilevel health services data: A Canadian neonatal mortality study. *Health Services & Outcomes Research Methodology*, 2004, **5**: 5 – 26.

Maule MM, Zuccolo L & Magnani C *et al*. Bayesian methods for early detection of changes in childhood cancer incidence: Trends for acute lymphoblastic leukaemia are consistent with an infectious aetiology. *European Journal of Cancer*, 2006, **42**: 78 - 83.

McCullagh P & Nelder JA. *Generalized Linear Models*, 2nd ed. London: Chapman and Hall, 1989.

Møller B. Prediction of cancer incidence — Methodological considerations and trends in the Nordic countries 1958 - 2022. Dissertation, Faculty of Medicine, University of Oslo, Unipub AS, Oslo, 2004.

Møller B, Fekjær H & Hakulinen T *et al*. Prediction of cancer incidence in the Nordic countries: empirical comparison of difference approaches. *Stat Med*, 2003, **22**: 2751 - 66.

Møller B, Fekjær H & Hakulinen T *et al*. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev*, 2002, **11** (Suppl): 1 - 96.

Møller H, Fairley L & Coupland V *et al*. The future burden of cancer in England: incidence and numbers of new patients in 2020. *British J Cancer*, 2007, **96**: 1484 - 8.

Osmond C. Using age, period and cohort models to estimate future mortality rates. *Inter J Epidem*, 1985, **14**: 124 - 9.

Qiu Z, Jiang Z & Hatcher J. Comparison of Short-Term Projection Methods: Proposing the Hybrid Approach and Bayesian Models. Technique Report for Cancer Projections Network (C-Proj), 2010.

Qiu Z, Song PXK & Tan M. Bayesian hierarchical models for multi-level repeated ordinal data using WinBUGS. *J Biopharma Stat*, 2002, **12**: 121 - 35.

Raftery AL & Lewis S. How many iterations in the Gibbs sampler? In Bernardo JM, Berger JO, Dawid AP & Smith AFM Eds. *Bayesian Statistics* 4. Oxford: Oxford University Press, 1992, pp763 – 773.

Ramsay RO & Silverman BW. *Functional Data Analysis*. New York: Springer, 1997.

Robert CP & Casella G. *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

Sharp L, Black RJ & Muir CS *et al*. Will the Scottish cancer target for the year 2000 be met? The use of cancer registration and death records to predict future cancer incidence and mortality in Scotland. *British Journal of Cancer*, 1996, 73: 1115 – 21.

Shibuya K, Inoue M & Lopez AD. Statistical modeling and projections of lung cancer mortality in 4 industrialized countries. *Int J Cancer*, 2005, **117**: 476 – 485.

Smirnov, NV. Tables for estimating the goodness of fit of empirical distributions, *Annals of Mathematical Statistic*, 1948, **19**: 279.

Stewart P & Xie L. Results of survey on cancer projection methods used in the provincial / territorial cancer registries. Presentation at Canadian Cancer Statistics Steering Committee Meeting, Toronto, June 23, 2009.

Tiwari RC, Ghosh K & Jemal A *et al*. A new method of predicting US and state-level cancer mortality counts for the current calendar year. *CA Cancer J Clin*, 2004, **54**: 30 - 40.

Verdecchia A, De Angelis G & Capocaccia R. Estimation and projections of cancer prevalence from cancer registry data. *Stat Med*, 2002, **21**: 3511 - 26.

Wingo PA, Landis S & Parker S *et al*. 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. *J Registry Management*, 1998, May: 43 - 51.

Wood SN. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL, USA: Chapman & Hall, 2006.

Xiao Y, Gordon A & Yakovlev A. A C++ program for the Cramer-von Mises two sample test, *Journal of Statistical Software*, 2007, **17**(8),

Xie L. Predication of cancer incidence and mortality in Canada. Presentation by Public Health Agency of Canada, 2010.

Yasmeen F, Hyndman RJ & Erbas B. Forecasting age-related changes in breast cancer mortality among white and black US women: A functional data approach. *Cancer Epidemiology*, 2010.