STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:2751–2766 (DOI: 10.1002/sim.1481)

Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches

Bjørn Møller[1,*,†], Harald Fekjær[1], Timo Hakulinen[2], Helgi Sigvaldason[3], Hans H. Storm[4], Mats Talbäck[5] and Tor Haldorsen[1]

[1]*Kreftregisteret, Institute of Population-based Cancer Research, N-0310 Oslo, Norway*
[2]*Finnish Cancer Registry, FIN-00170 Helsinki, Finland*
[3]*Icelandic Cancer Registry, IS-125 Reykjavik, Iceland*
[4]*Danish Cancer Society, DK-2100 Copenhagen, Denmark*
[5]*Centre for Epidemiology, The National Board of Health and Welfare, S-106 30 Stockholm, Sweden*

SUMMARY

Prediction of the future number of cancer cases is of great interest to society. The classical approach is to use the age-period-cohort model for making cancer incidence predictions. We made an empirical comparison of different versions of this model, using data from cancer registries in the Nordic countries for the period 1958–1997. We have applied 15 different methods to 20 sites for each sex in Denmark, Finland, Norway and Sweden. Median absolute value of the relative difference between observed and predicted numbers of cases for these 160 combinations of site, sex and country was calculated. The medians varied between 10.4 per cent and 15.3 per cent in predictions 10 years ahead, and between 15.1 per cent and 32.0 per cent for 20 year predictions. We have four main conclusions: (i) projecting current trends worked better than assuming that future rates are equal to present rates; (ii) the method based on the multiplicative APC model often overestimated the number of cancer cases due to its exponential growth over time, but using a power function to level off this growth improved the predictions; (iii) projecting only half of the trend after the first 10 years also gave better long-term predictions; (iv) methods that emphasize trends in the last decade seem to perform better than those that include earlier time trends. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: cancer incidence; prediction; projection; forecast; age-period-cohort model; evaluation

1. INTRODUCTION

There is in society a general desire for predictions of the future cancer burden, but it is always difficult to foresee the future. Still, making predictions about future cancer incidence is an interesting subject, which includes statistical modelling, model fitting and interpretation. At the same time, it is a revealing task, because the truth will sooner or later catch up with you and judge your performance.

* Correspondence to: Bjørn Møller, Kreftregisteret, Institute of Population-based Cancer Research, Montebello, N-0310 Oslo, Norway.
† E-mail: bjorn.moller@kreftregisteret.no

Contract/grant sponsor: Nordic Cancer Union.

Received January 2002
Accepted December 2002

There are various administrative and scientific reasons for predicting future cancer incidence [1]. Estimates of the size of the future cancer burden are very important for health planning, and predictions have been used to allocate appropriate amounts of resources for diagnosis, treatment and rehabilitation [2]. The question of how well prediction models perform, however, depends on what we are aiming at. If we want to evaluate the effect of some type of intervention, such as a screening programme for *in situ* cervix cancer, the predicted rate of cervix cancer can be judged by its contrast to the future observed rates [3]. However, we are primarily interested in having predictions that are accurate. Thus, the different prediction methods will be judged by their ability to produce predictions that come as close as possible to what is observed in the future.

Predictions of cancer incidence have previously been made for all the Nordic countries up to the year 2010, based on observed rates up to 1987 [4]. The aim of this study is to evaluate how successful these predictions have been, and to compare the methods used by Engeland *et al*. [4] with alternative methods.

2. MATERIAL

The material consists of new cancer cases reported to cancer registries in Denmark, Finland, Iceland, Norway and Sweden between 1958 and 1997, with population figures covering the same calendar period from the central statistical offices in the different countries. The Nordic cancer registries receive reports from physicians, hospitals, pathological and cytological laboratories, and (except in Sweden) from death certificates [5]. Compulsory reporting, combined with reporting from multiple sources, ensures almost 100 per cent coverage of all the cancer cases [6].

Table I lists the types of cancer sites included in the study. A detailed description of the types of tumours that are included at each site is given elsewhere [4]. Owing to the small number of cases, lip cancer and larynx cancer for women and breast cancer for men were included in 'other sites'. Cancer of the breast, cervix uteri and prostate have been subjected to early detection programmes or practices. These three sites will collectively be referred to as intervention sites, and figures will be given with and without these sites when appropriate.

The data is tabulated in five-year age groups $(0–4, 5–9, \ldots, 80–84, 85+)$ and calendar periods, either five-year (1958–1962, 1963–1967, $\ldots$, 1993–1997) or one-year (1958, 1959, $\ldots$, 1997). In Denmark, cases are not available for 1997, so average numbers of cases in 1993–1996 are used. Since there are 20 types of cancers for each sex, with five countries in the study, we have 200 different combinations to make predictions for. In Engeland *et al.* [4], Iceland was given special treatment due to the small number of cases there. Because of this, Iceland is included when evaluating the previous predictions, but excluded when comparing the different prediction methods.

3. CRITERIA FOR COMPARISON

Engeland *et al.* [4] made predictions for 1993–1997, based on observed rates up to 1987. This will be referred to as recent short term predictions, see Figure 1. To be able to study the consistency of different prediction methods, we also include predictions for the previous

Table I. Types of cancer included in the study.

Type of cancer	ICD-7	Males	Females
Lip	140	x	
Tongue, oral cavity and pharynx	141, 143–145, 147, 148	x	x
Oesophagus	150	x	x
Stomach	151	x	x
Colon	153	x	x
Rectum	154	x	x
Pancreas	157	x	x
Larynx	161	x	
Lung	162–163	x	x
Breast	170		x
Cervix uteri	171		x
Corpus uteri	172		x
Ovary	175.0		x
Prostate	177	x	
Testis	178	x	
Kidney	180	x	x
Urinary bladder	181.0	x	x
Melanoma of the skin	190	x	x
Thyroid	194	x	x
Non-Hodgkin's lymphomas	200, 202	x	x
Hodgkin's disease	201	x	x
Multiple myeloma	203	x	x
Acute leukaemia	204 AL	x	x
Other sites	—	x	x

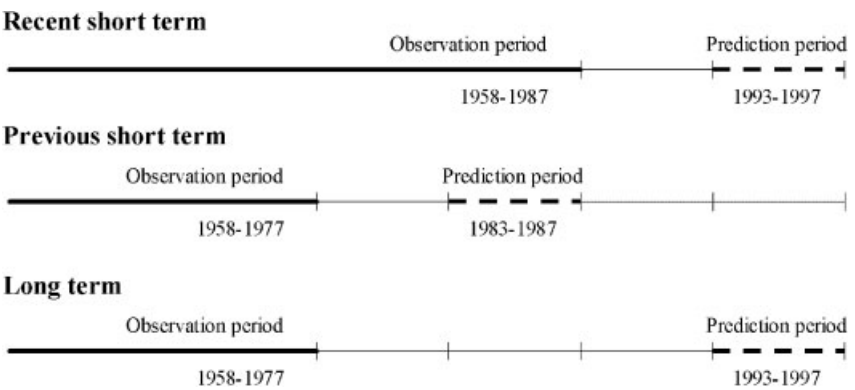

Figure 1. Observation base for different length of predictions.

period, 1983–1987, based on observed rates up to 1977. Both long term and short term predictions are considered important, and since we only have observed rates up to 1997, long term predictions are based on observed rates up to 1977.

The distance between observed and predicted values for all age groups combined, is measured by the absolute value of the relative difference between observed and predicted number

of cases: |observed – predicted|/observed. Excluding Iceland, there are 160 combinations of site, sex and country. The median of the 160 distances is used as a summary statistic. When comparing each separate age group, the absolute difference is used, because young age groups, with relatively fewer cases, would dominate unduly if relative difference had been used. The median is calculated from the 2880 combinations of 20 sites, two sexes, four countries and 18 age groups. *A priori*, predictions for all age groups combined are decided to be more important than age-specific estimates, especially regarding health planning, where the age of a patient is secondary. However, age specific rates should be taken into account when different methods perform similarly with respect to all age groups combined.

Even if we knew the true underlying rates in the future, there would still be a difference between the expected and the observed numbers of cases, due to the inherent random variation. This difference can be viewed as a limit of how close observed and predicted values theoretically can get. If we assume that the number of cases follow a Poisson distribution, this difference can be calculated by simulation.

Friedman's test [7] is used to test for statistical difference in medians between different prediction methods. The test is non-parametric, and is based on internal ranks between the methods. Each combination of site, sex and country is considered to be one separate block. Pearson's correlation coefficient [7] is used to calculate the association between the performance of various prediction methods in different periods.

4. PREDICTION METHODS

Historically, the age-period-cohort (APC) model [8] has been widely used for making cancer incidence predictions. This Poisson regression model was also used in the previous cancer incidence predictions in the Nordic countries, but with a couple of changes based on earlier experience [9]. To level off the exponential growth in the multiplicative model, a power link was used instead of the log link. Secondly, only half the linear trend was projected for the third and fourth five-year prediction period, based on the belief that trends would eventually fade off.

While the previously applied method used five-year periods for the last 30 years, we will also include two alternative methods which use the last 11 one-year periods. One is based on simple linear [10] and non-linear [11] models and one on a smoothed version of the APC model [12]. By fitting different versions of the models, we shall evaluate the effect of: (i) altering the link between the rate and the covariates; (ii) cutting the trend component; (iii) emphasizing recent trends.

Common for all methods is the assumption that current trends, or parts of them, will continue into the future. To evaluate whether this assumption can be justified, we shall also compare the trend models with a basic model, which assumes that the rates are constant. All methods are labelled with a letter indicating which main method they originate from, and a number separating the different modifications.

For sites with few cases in young age groups, average rates from the last 10 years were used in the predictions for those age groups, see Engeland *et al.* [4] for details. In practice, all predictions must be based on forecasted population numbers, but since our main focus is on changes in cancer trends, not population changes, all calculations are based on observed population figures.

The analyses were done in S-plus and R, using the functions `glm()` or `ms()` to estimate the parameters in the different models.

4.1. Methods A0 and A1 (The APC model)

Method A0 is based on the classical age-period-cohort model, which has often been used for prediction of cancer incidence and mortality [13–17]. It is based on tables with five-year age groups and five-year calendar periods. Birth cohort is constructed synthetically by subtracting age from period. The model can be written as $R_{ap} = \exp(A_a + Dp + P_p + C_c)$, where R_{ap} is the incidence rate in age group a in calendar period p, D is the common drift parameter [18], A_a is the age component for age group a, P_p is the non-linear period component of period p and C_c is the non-linear cohort component of cohort c.

Future non-linear cohort and period effects are put equal to the last estimated effect in the model. Predictions are made by projecting the drift, but instead of adding D to each of the four new periods, D, D, $1/2D$ and $1/2D$ added. This reduction of the drift was implemented in the previous predictions, but with a different link function in the model.

The APC model gives exponential growth in the predictions. One way to modify this is to plug the estimates from the model in method A0 into the formula $R_{ap} = \exp(A_a + P_p + C_c)(1 + Dp)$ in the predictions of cancer sites with increasing trends (that is, $\hat{D} > 0$). This is motivated by the fact that $(1 + x)$ is the first-order Taylor approximation of $\exp(x)$ and will level off the exponential growth. This method is named A1.

4.2. Methods B0, B1, B2, B3, B4 and B5 (The power model)

This is the approach adapted by Engeland *et al.* [4] in the previous predictions. Method B0 is based on a model that uses a power function instead of a logarithmic function as the link between the rate and the covariates [19]. Mathematically, the model is written $R_{ap} = (A_a + Dp + P_p + C_c)^5$, where R_{ap}, A_a, P_p and C_c are defined as in method A0.

Predictions are made by projecting the drift, with only half the drift in the third and fourth periods, as in method A0.

Method B0 is modified in several ways. One way is to not cut any drift, that is to add D in all the four future five-year periods (method B1). This will only influence long term predictions. To evaluate the effect of the non-linear cohort and period components, we include only age and drift in a model (method B2), in which the rate can be written as $R_{ap} = (A_a + Dp)^5$. The power of five in method B0 was chosen empirically to level off the exponential growth. Choosing a lower power will level off this growth even more, but with too small values, numerical problems will occur. In method B3, a power of two is used.

One might reason that recent periods are more relevant than earlier periods. This can be implemented in the model by weighting the cases (method B4). The most recent period is given w times more weight than the earliest period, the second most recent period $(w - 1)$ times more weight etc., where w is the number of periods in the prediction base.

The drift component is the average trend for the whole period of observation. One way to make the method more dynamic is to use only the slope between the two most recent periods. If we have P periods, then the last slope is $D - P_{P-1} = D_{\text{last}}$, where D is the common drift and P_{P-1} is the last deviation from this. For smaller sites, the variance of D_{last} is large, which will give unreliable predictions. This means that we are only interested in using D_{last} if there are changes in the trend over time that are not caused by chance. The model

$R_{ap} = (A_a + Dp + Sp^2 + C_c)^5$ is used to determine whether to apply D or D_{last} in the projection of the common drift (method B5). If there is a significant curvature in the trend over time, that is, if $\hat{S}$ is significant, D_{last} is used, otherwise we use D.

4.3. Methods C0, C1 and C2 (The short-base model)

Method C0 uses only the last 11 one-year periods. It was used to make predictions up to the year 2005 for five oncological regions in Finland [20]. Four different submodels are used:

$$\begin{array}{ll} \text{(I)} \quad R_{ap} = A_a(1 + Dp) & \text{(II)} \quad R_{ap} = A_a + D_a p \\ \text{(III)} \quad R_{ap} = \exp(A_a + Dp) & \text{(IV)} \quad R_{ap} = \exp(A_a + D_a p) \end{array}$$

where A_a is the age component of age group a, D is the common drift parameter for all age groups, and D_a is the drift parameter corresponding to age group a. Models (I) and (II), both linear in time, are used for cancer sites with an increasing trend, whereas the log-linear models, (III) and (IV), are used for sites with a decreasing trend. The sign of the drift-parameter D in model (III) is used to classify the specific site as increasing or decreasing. Model (I), which is non-linear in parameters, can be written on the same form as model (II), with the restriction $D_a = A_a D$. Models (II) or (IV), with separate slopes for each age group, are used if they give a significant improvement compared to models (I) or (III), respectively. Reduction in deviance is tested at the 5 per cent level.

To make predictions comparable to methods A0 and B0, three times D is added for the first five-year prediction period. This is equal to adding D for each year for the next five years and taking the average of these to calculate the predicted rate for the five-year period. Only half the drift is added in the third and fourth prediction periods.

Model (I), which is non-linear in parameters, is not implemented in most standard statistical software packages. This encourages a modification of the method. Instead of estimating the parameters in (I) and (II) for cancer sites with increasing trends, we can create a version, method C1, by estimating parameters in (III) and (IV). These estimates can be plugged into the formulae of (I) and (II), respectively. Because $(1 + x) \approx \exp(x)$ for small values of x, this will only marginally change the predictions. In addition, age groups with increasing trends for a cancer site with an overall decreasing trend will also be projected using a linear and not a log-linear formula, thus avoiding exponential growth in the future rates.

Figure 2 shows the motivation for the next modification. Let A be the number of age groups in the model. Instead of choosing between one drift parameter as in (III) and A drift parameters as in (IV), we add the number of polynomials that is sufficient to describe the slopes in the different age groups. In Figure 2 we see that a linear term adequately describes the different age specific slopes in female stomach cancer in Sweden. This is the model $R_{ap} = \exp(A_a + D_{ac} p)$, where $D_{ac} = D_0 + D_1 a$. Here $c = p - a$, where c is the synthetic cohort group corresponding to period $p = 1, 2, \ldots, 11$ and age group $a = 5, 10, \ldots, 5A$. We have included c in the notation because it might be that different slopes for different cohort groups would give a better fit. Whether to expand with polynomials of age or cohort is decided by comparing the two models with $D_{ac} = D_0 + D_1 a$ and $D_{ac} = D_0 + D_1 c$. Assume that the model with slopes linear in age gives less deviance of the two. The number of age polynomials to include in the final model will then be chosen in the following way. The first step is to compare the models with $D_{ac} = D_0$ and $D_{ac} = D_0 + D_1 a$. If $\hat{D}_1$ is significantly different from zero, continue to add polynomials $D_2 a^2$, $D_3 a^3$ etc., until both $\hat{D}_k$ and $\hat{D}_{k+1}$ are non-significant. If the model with slopes linear

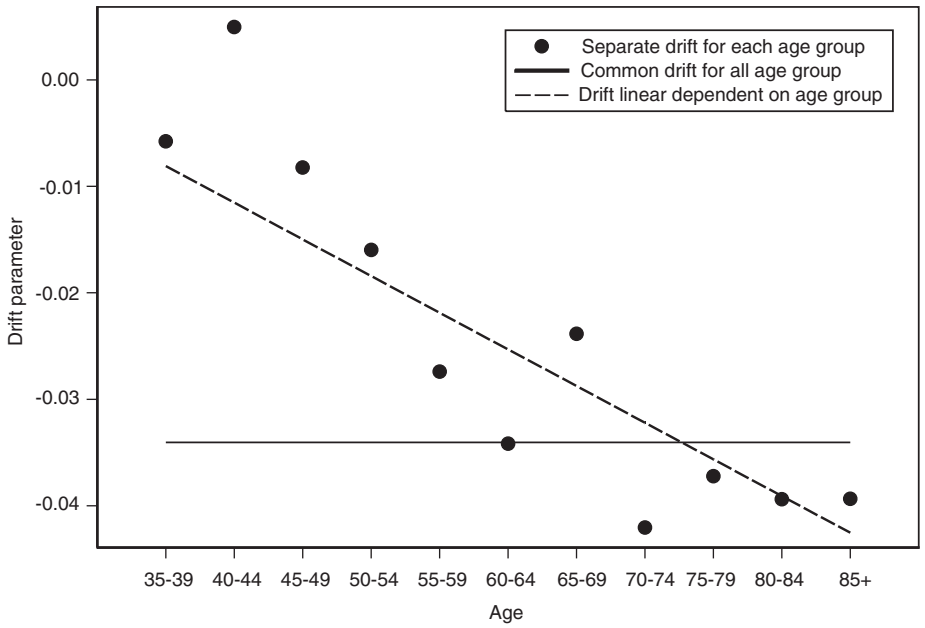

Figure 2. Drift parameters for different age groups. Female stomach cancer in Sweden, recent short term predictions.

in cohort gives less deviance, the final model is chosen in a similar way, but by adding polynomials of cohort instead of age. For age groups with increasing trends (that is, $\hat{D}_{ac} > 0$) estimates are plugged into the formula $R_{ap} = \exp(A_a)(1 + D_{ac}\, p)$. This method is named C2.

4.4. Methods D0 and D1 (The short APC model)

A full APC model applied on one-year rates would include 18 age groups, 11 periods and 28 $(18 + 11 - 1)$ cohorts. This motivates some sort of smoothing on the non-linear period and cohort effects. We include second- and third-order polynomials as a way to smooth the effects. The D0 method uses the model $R_{ap} = \exp(A_a + Dp + P_2 p^2 + P_3 p^3 + C_2 c^2 + C_3 c^3)$, where D is the common drift parameter, P_2 and P_3 are the parameters for the second- and third-order polynomial of period, C_2 and C_3 are the second- and third-order polynomial of cohort, and p and c are the centred values for period and cohort, respectively. This model is along the lines of what was used in Coleman *et al.* [12], but with a fixed number of polynomials instead of an algorithm for selecting the optimal order.

The polynomials in the model are only included to smooth the non-linear period and cohort effects. In the predictions, new cohorts are given the value of the youngest observed cohort, new periods are given the value of the last observed period, and only the common drift parameter is projected.

A problem with the smoothed APC model applied on one-year calendar periods is that the polynomials used for smoothing the non-linear period effects can have large variances. We can create a modified version, method D1, by applying the model to the last 16 one-year

rates. If $\hat{P}_2$ in the model $R_{ap} = \exp(A_a + Dp + P_2 p^2 + C_2 c^2 + C_3 c^3)$ is significant, we use the 11-year base, otherwise we use the 16-year base.

4.5. Methods E0 and E1 (Present state model)

A natural baseline to use as a comparison for all the above-mentioned methods is to assume that the age specific rates in the future will be the same as the last observed rates. To get more stable rates, we use the average of the last five years for each age group: $R_{ap} = \exp(A_a)$.

These rates can be used together with population forecasts to calculate the predicted number of cases in the future. An even simpler approach, method E1, is to use the average observed number of cases in the previous 5 years as our estimate of the number of cases in future five year periods. All the 15 different methods are summarized in Table II.

5. RESULTS

5.1. Evaluation of the power model predictions

For all of the five Nordic countries combined, incidence rates increased by 4.6 per cent from 1983–1987 to 1993–1997, whereas Engeland *et al.* [4] predicted a 7.6 per cent increase (method B0). Looking at all 200 combinations of country, sex and site separately, 53.8 per cent had an increasing trend, while 65.6 per cent were predicted to increase. When excluding the intervention sites, the average increase is reduced to 2.6 per cent, compared to a predicted increase of 7.8 per cent. Two of the intervention sites, cancers of the breast and prostate, increased much more than predicted from 1983–1987 to 1993–1997, due to the introduction of mammographic screening for breast cancer and the availability of testing for prostate-specific antigen (PSA) for prostate cancer. For further details, see Møller *et al.* [21].

The boxes in Figure 3 contain the first and third quartile, of the absolute value of the relative difference between observed and predicted numbers of cases using the power model, from all the 40 combinations of sex and site. The performance in each country is rather similar. The exception is Iceland, for which the predictions were somewhat more erratic.

The power model was used in the previous predictions to level off the exponential growth of the multiplicative APC model. Table III shows that this choice improved the predictions; the median of absolute value of the relative difference between observed and predicted numbers of cases are lower for B0 than for A0, both in the short and long term predictions.

The second attempt to improve predictions was to project only half the trend after the first 10 years. This modification is only relevant for long term predictions. The median difference is 21.8 per cent if the drift is projected for 20 years (B1), whereas it reduces to 19.5 per cent when cutting half the drift in the last 10 years (B0). Similar effects are seen in Table IV for age specific predictions. In Figure 4 the effect of using a power link and cutting the drift is visualized. Looking at the first 10 years of prediction using the APC model, we see that the rates grow exponentially. The break in the curve after 10 years is the effect of cutting the drift component of the model.

The median difference for the five countries combined is 13.4 per cent. If we knew the true underlying rates in the future, the median of the absolute value of the relative difference is found by simulation to be 1.6 per cent. This gives the lower limit due to pure Poisson variation, and shows that Poisson variations do not represent the major part of the discrepancy

Table II. Overview of the different prediction methods.

Method names	Estimation issues						Prediction issues			
	Base (years)	Weighting	Non-linear period and cohort terms	Nonlinear in parameters	Age/cohort dependent slope	Fitted model for the rate $(R_{ap})^*$	Projection formula for the rate $(R_{ap})^*$	Recent slope	Halving of drift	Linearization of drift
APC (A0)	20–30	No	Yes	No	No	$\exp(A_a + Dp + P_p + C_c)$	$\exp(A_a + Dp + P_p + C_c)$	No	Yes	No
Plug-in APC (A1)	20–30	No	Yes	No	No	$\exp(A_a + Dp + P_p + C_c)$	$\exp(A_a + P_p + C_c)(1 + Dp)$	No	Yes	Yes
Power (B0)	20–30	No	Yes	No	No	$(A_a + Dp + P_p + C_c)^5$	$(A_a + Dp + P_p + C_c)^5$	No	Yes	No
Full drift (B1)	20–30	No	Yes	No	No	$(A_a + Dp + P_p + C_c)^5$	$(A_a + Dp + P_p + C_c)^5$	No	No	No
Age + drift (B2)	20–30	No	No	No	No	$(A_a + Dp)^5$	$(A_a + Dp)^5$	No	Yes	No
Power of two (B3)	20–30	No	Yes	No	No	$(A_a + Dp + P_p + C_c)^2$	$(A_a + Dp + P_p + C_c)^2$	No	Yes	No
Time weigh. (B4)	20–30	Yes	Yes	No	No	$(A_a + Dp + P_p + C_c)^5$	$(A_a + Dp + P_p + C_c)^5$	No	Yes	No
Recent slope (B5)	20–30	No	Yes	No	No	$(A_a + Dp + P_p + C_c)^5$	$(A_a + Dp + P_p + C_c)^5$	Yes	Yes	No
Short-base (C0)	11	No	No	Yes	No	↗: $A_a(1 + Dp)$ or $A_a + D_a p$ ↘: $\exp(A_a + Dp)$ or $\exp(A_a + D_a p)$	↗: $A_a(1 + Dp)$ or $A_a + D_a p$ ↘: $\exp(A_a + Dp)$ or $\exp(A_a + D_a p)$	No	Yes	No
Plug-in short-base (C1)	11	No	No	No	No	↗: $\exp(A_a + Dp)$ or $\exp(A_a + D_a p)$ ↘: $\exp(A_a + Dp)$ or $\exp(A_a + D_a p)$	↗: $\exp(A_a)(1 + Dp)$ or $\exp(A_a)(1 + D_a p)$ ↘: $\exp(A_a + Dp)$ or $\exp(A_a + D_a p)$	No	Yes	Yes
AC-slope (C2)	11	No	No	No	Yes	↗: $\exp(A_a + D_{ac} p)$ ↘: $\exp(A_a + D_{ac} p)$	↗: $\exp(A_a)(1 + D_{ac} p)$ ↘: $\exp(A_a + D_{ac} p)$	No	Yes	Yes
Short-APC (D0)	11	No	Yes	No	No	$\exp(A_a + Dp + P_2 p^2 + P_3 p^3 + C_2 c^2 + C_3 c^3)$	$\exp(A_a + P_2 p^2 + P_3 p^3 + C_2 c^2 + C_3 c^3)(1 + Dp)$	No	Yes	Yes
16-year base (D1)	11–16	No	Yes	No	No	$\exp(A_a + Dp + P_2 p^2 + P_3 p^3 + C_2 c^2 + C_3 c^3)$	$\exp(A_a + P_2 p^2 + P_3 p^3 + C_2 c^2 + C_3 c^3)(1 + Dp)$	No	Yes	Yes
Present state (E0)	5	No	No	No	No	$\exp(A_a)$	$\exp(A_a)$	No	—	—
Naive (E1)	5	No	No	No	No	# cases	# cases	No	—	—

*For methods where different models are applied for cancer sites with increasing and decreasing trends, ↗indicates sites with increasing trend and ↘ indicate sites with decreasing trends.

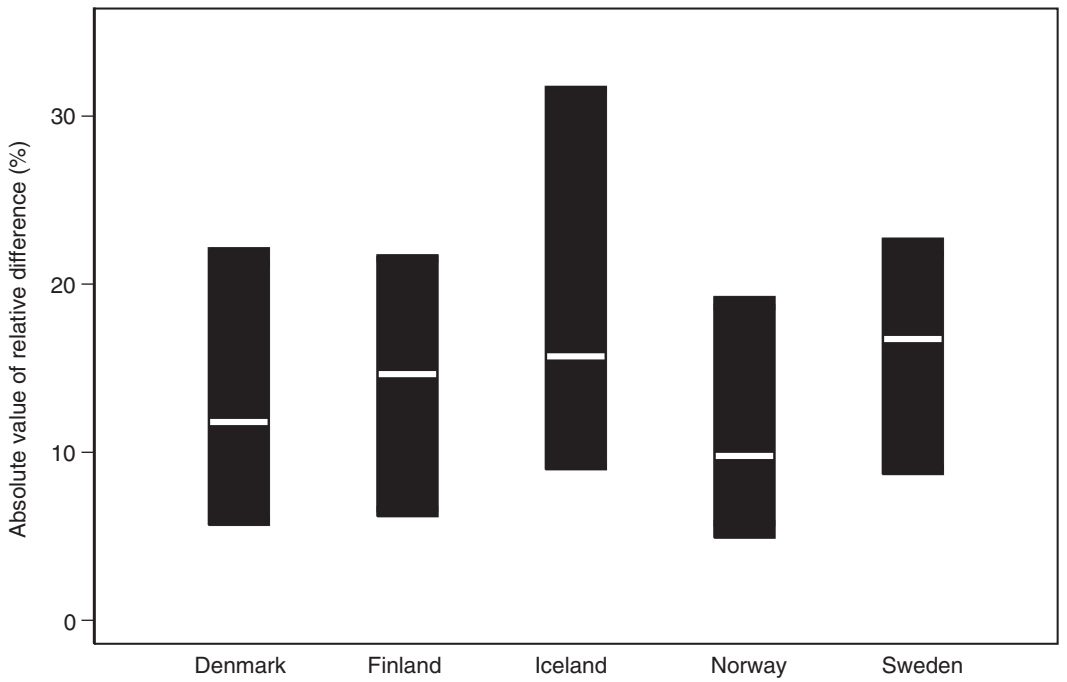

Figure 3. Box plot of absolute values of the relative differences between observed number and predicted number of cases in 1993–1997, based on 10 year predictions using the power model, by country.

Table III. All age groups combined. Median of absolute value of the relative difference between observed and predicted number of cases, by type of prediction method and length of prediction (per cent). Median of all 160 combinations of site, sex, country. Iceland is excluded.

Prediction method	Length of prediction		
	Previous short term (10 years)	Recent short term (10 years)	Long term (20 years)
APC (method A0)	11.8	14.2	22.0
Plug-in APC (A1)	9.7	13.5	17.0
Power (method B0)	9.4	13.1	19.5
Full drift (B1)	9.4	13.1	21.8
Age+drift only (B2)	12.9	12.7	24.1
Power of two (B3)	8.9	11.8	20.1
Time weighted (B4)	9.4	13.0	19.3
Recent slope (B5)	9.3	12.1	17.6
Short-base (method C0)	9.3	12.5	15.8
Plug-in short-base (C1)	8.8	11.9	15.1
AC-slope (C2)	9.4	10.4	18.1
Short-APC (method D0)	11.5	12.4	20.1
16-year base (D1)	11.4	11.8	20.1
Present state (method E0)	15.0	14.4	23.7
Naive (E1)	19.9	15.3	32.0

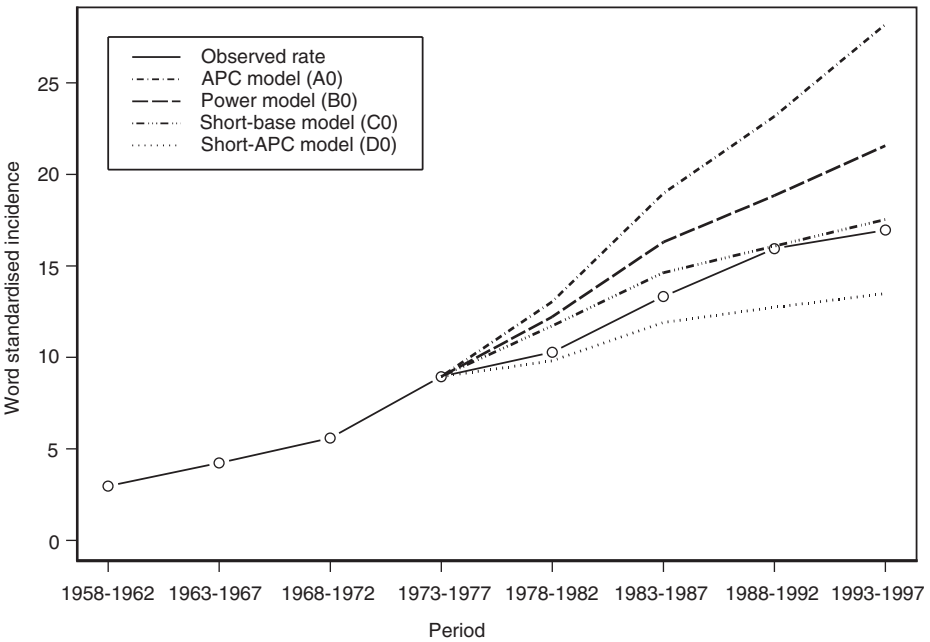

Figure 4. Observed versus predicted age-adjusted incidence rate of melanoma of the skin for males in Norway. Predicted rates are based on observations up to 1977 for the four main prediction methods.

Table IV. Age-specific predictions. Median of the absolute value of the difference between observed and predicted number of cases, by type of prediction method and length of prediction. Median of all 2880 combinations of site, sex, country and age group. Iceland is excluded.

Prediction method	Length of prediction		
	Previous short term (10 years)	Recent short term (10 years)	Long term (20 years)
APC (method A0)	19.0	21.4	34.0
Plug-in APC (A1)	17.2	20.2	30.8
Power (method B0)	16.2	19.4	31.1
Full drift (B1)	16.2	19.4	35.6
Age+drift only (B2)	20.3	23.2	36.1
Power of two (B3)	16.9	17.7	34.0
Time weighted (B4)	17.1	19.2	31.8
Recent slope (B5)	16.2	19.3	30.9
Short-base (method C0)	19.7	21.7	31.3
Plug-in Short-base (C1)	17.8	20.6	29.3
AC-slope (C2)	17.7	19.1	30.5
Short-APC (method D0)	18.5	20.2	31.7
16-year base (D1)	17.9	18.5	30.6
Present state (method E0)	19.5	20.6	35.1
Naive (E1)	26.5	23.0	40.0

between observed and projected numbers of cases. The gap between this lower limit and the median difference based on the applied prediction method implies that there is a large potential for improvements in modelling.

5.2. Comparison of different prediction methods

Medians of absolute value of relative difference between an observed number of cases, and the number of cases predicted by each of the different methods, are given in Table III. Looking at the four main methods A0–D0, the average median difference for previous short term, recent short term and long term are 10.5 per cent, 13.0 per cent and 19.4 per cent, respectively. When using the present state model, the figures are 15.0 per cent, 14.4 per cent and 23.7 per cent, respectively. Three of the main methods, using the power, short-base and short APC model, are all significantly better than the present state model, both for the long term prediction and for the two short term predictions combined ($p < 0.02$ for all pairwise comparisons). The APC model is not significantly better ($p = 0.58$ and $p = 0.34$ for short term and long term predictions, respectively). The four main methods have overall significantly different medians, both for long term ($p < 0.01$) and short term predictions ($p < 0.01$). The short-base model performs marginally better for all age groups combined, especially in the long term predictions.

Looking at predictions for separate age groups in Table IV, the power model has the lowest median of absolute difference between observed and predicted numbers of cases. The gap to the other methods is most visible for the two short term predictions. The other main methods, A0, C0 and D0, had problems outperforming the present state model (E0).

In Tables III and IV, we can see the effects of the various modifications of the methods. Of the four main methods, the classical APC model (A0) gives the worst performance. Plugging the estimates of the model into a linear formula for sites with increasing trends (A1) improves the method both for short term and long term predictions. However, the improvements are not large enough to outperform any of the other main methods.

The version of the power model with only age components and a common drift parameter (B2) does not perform as well as the main method (B0), which includes non-linear period and cohort components. The pattern is evident, both for all age groups combined and for age-specific predictions. Using the power of two (B3) instead of the power of five as in the main method (B0) seems to improve the short term predictions, while this is not the case for long term predictions. One problem when reducing the power is that the fit of the model usually declines as a function of the power.

Giving more weight to recent periods has little effect on the performance of the method. The time weighted model (B4) gives approximately the same predictions as the unweighted model (B0). The last modification of the power model is to use the most recent slope as the drift component in the projections (B5). This seems to improve the predictions for all age groups combined.

In the short-base model (C0), the non-linear model for increasing trends is estimated directly. Alternatively, we estimate the parameters in a multiplicative model and plug the estimates into a linear formula (C1). This results in slightly better predictions compared to the main method (C0). The model with slopes potentially dependent on age or cohort (C2) performs in line with the plug-in version of the short-base model (C1).

The smoothed APC model with 11-year base (D0) produced mediocre results. Using a 16-year base where this is appropriate (D1) improves the performance only marginally.

To investigate consistency of the 15 different methods over time, correlation of median performance between previous and recent short term predictions is calculated. For all age groups combined, the correlation is 0.49 (95 per cent CI −0.05–0.81). Excluding the present state model (E0), the correlation drops to 0.28. The correlation is much higher for age specific predictions, 0.81 (95 per cent CI 0.49–0.94), excluding the present state model.

6. DISCUSSION

Prediction of cancer incidence in the Nordic countries [4] was based on the classical age-period-cohort model, with a couple of modifications. A power function was used instead of the logarithm as the link function, and the projected trends were halved after the first 10 years of prediction. Both of these changes have been found in our study to improve the predictions.

Fifteen prediction methods have been compared, based on eight (A0, B0, B2, B3, C0, C2, D0 and E0) different statistical models, which were fitted to 320 different data sets (2 sexes $\times$ 20 sites $\times$ 4 countries $\times$ 2 periods). A sharp analysis, including residual plots of each data set, was thus unpractical. Model validation, measured as the distance between the observed number of cases in the future and the predicted values, has been our main criterion when comparing the methods. How well the models fit has not been a criterion, but we have made an effort to make sure that none of the models gave poor fit. For methods using models with a long prediction base (power model and APC model), the model fit was ensured by successively excluding the earliest calendar periods when goodness-of-fit of the model was rejected. For the short-base model, goodness of fit was handled by modelling separate slopes for each age group when common slope gave a poor fit. In the short APC model, we chose to use a fixed number of parameters to smooth the non-linear effects of cohort and period. The prediction base of 11 years was short enough to give an acceptable fit for most sites.

Alternative methods to smooth the effects in the APC model have been suggested. Coleman *et al.* [12] used polynomials, as we did, but chose the number of polynomials based on goodness-of-fit of the model. Berzuini and Clayton [22] have a Bayesian approach, using autoregressive priors to smooth the effects. Common to both approaches is that they give predictions with exponential growth, which for some sites can give extreme predictions [23].

There is not much literature on evaluation of previous predictions. One exception is predictions done in Finland for the year 1980, based on observed rates up to 1968 [24] using linear and log-linear models. The predictions were evaluated by calculating 90 per cent prediction intervals [25]. In our study, we have made a direct comparison between point estimates and observed number of cases, a method that is more suitable when the performance of different prediction methods is compared. Another exception is an empirical evaluation of the Bayesian age-period-cohort model [26], based on selected sites in Finland. The Bayesian approach was found to work better than projecting trends for each age group separately, but an age-cohort model, based on maximum likelihood estimates, preformed equally well.

Are methods that performed favourably in previous time periods guaranteed to perform well in the future? This might be answered by looking at how consistently the different methods

perform over time. For all age groups combined, we found a non-significant correlation between the previous short term performance and the recent short term performance of 0.49. The methods were more consistent for age-specific predictions, with a corresponding correlation of 0.81. Despite large variations, these correlations indicate that methods performing well in one time period also perform well in another.

Nearly all the methods examined in the present study performed better than the present rates as predictors for future rates. However, the relative gain by using linear models depends on the time periods in which the models are applied. Using the present state model for short term predictions, the median of the absolute value of the relative difference between observed number and predicted number of cases was about 15 per cent both for 1983–1987 and for 1993–1997. Using different types of linear models to predict the number of cases for the same periods, the median difference was about 9–12 per cent in the earliest period, while it was as high as 11–14 per cent in the most recent period. This shows that trends continued to a lesser extent after 1987 than they did 10 years earlier.

In addition to evaluating the effect of different choices made in the previous predictions, several other methods and modifications were considered. All the four main methods worked rather similarly, but the short-base model performed better for long term prediction. The reason why this method did well on long term predictions is probably that the projected drift of the model used is linear over time, not multiplicative as in the APC model, or close to multiplicative as in the power model. This is further supported by the fact that the short-base model overestimated the rates in 59 per cent of the long term predictions, while the APC model overestimated in 72 per cent.

In the plug-in version of the short-base model, the parameters were estimated using a multiplicative model, but for age groups with increasing trends these estimates were plugged into a linear formula. It might be questioned whether it is appropriate to use one model to estimate parameters and a different model to make the predictions. We will, however, argue that the plug-in method was constructed to level off the exponential drift. Since $\exp(Dp)$ is always greater than its first-order approximation $(1+Dp)$, the plug-in method gives more conservative predictions, that is, closer to present rates. A drawback of the short-base model is that it does not include cohort-specific information. A method using a modified model, the AC-slope model, in which the slopes for the different age groups were potentially dependent on cohort or age, seemed to work well.

The length of the prediction base is important. If we choose a long base, we utilize more information, but we increase the risk of including trends that have ceased to exist. Our results indicate that long base models suffer by including outdated information in the drift component. The modified version of the power model, called the recent slope method, which uses only the slope from the last 10 years as the drift in the projections, seemed to solve this problem. The recent slope method improved the predictions both for short and long term predictions. The two modifications of the short-base model both produced predictions in line with the recent slope method. Another way of explaining why these three methods tend to perform better is that if there is a curved trend, which continues into the future, it is better to be close to a recent tangent that to an older secant, see Figure 5. The age-period-cohort model is the traditional method for making cancer incidence predictions. The recent slope method is a natural modification of this model, so we recommend this method in practice. Software (in R and S-plus) for making predictions using the recent slope method, including power link and reduction of drift, is available at www.kreftregisteret.no/software/nordpred.

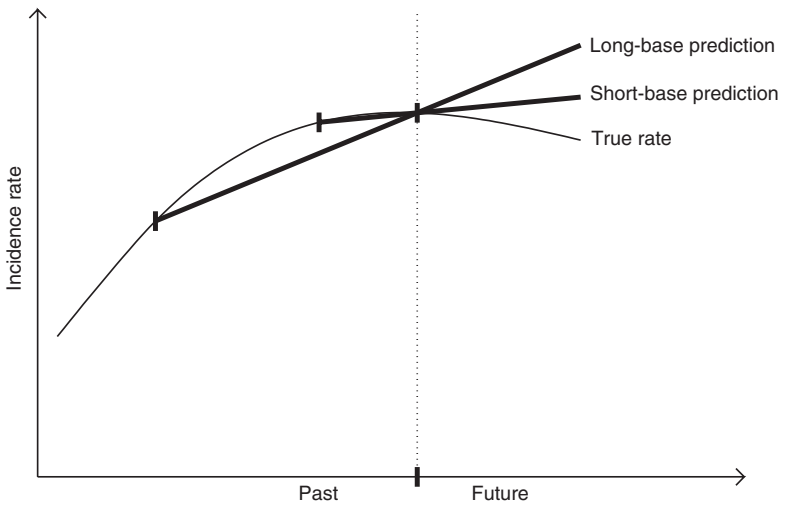

Figure 5. Predictions based on a short prediction base (recent tangent) versus long prediction base (old secant).

ACKNOWLEDGEMENTS

The authors would like to thank Petter Laake for valuable comments on the article. The project is supported by the Nordic Cancer Union.

REFERENCES

1. Hakulinen T, Hakama M. Predictions of epidemiology and the evaluation of cancer control measures and the setting of policy priorities. *Social Science & Medicine* 1991; **33**:1379–1383.
2. NOU Norsk kreftplan. Norwegian official report—National cancer plan for Norway, 1997; **20** (in Norwegian).
3. Beral V, Booth M. Predictions of cervical cancer incidence and mortality in England and Wales. *Lancet* 1986; **1**:495.
4. Engeland A, Haldorsen T, Tretli S, Hakulinen T, Hörte T, Luostarinen T, Magnus K, Schou G, Sigvaldason H, Storm HH, Tulinius H, Vaittinen P. Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. A collaborative study of the five Nordic Cancer Registries. *APMIS Supplementum* 1993; **38**:1–124.
5. Survey of Nordic Cancer Registries. At http://ncu.cancer.dk/ancr/survey.htm 2000. 3-21-2003.
6. Hakulinen T, Andersen A, Malker B, Pukkala E, Schou G, Tulinius H. Trends in cancer incidence in the Nordic countries. A collaborative study of the five Nordic Cancer Registries. *Acta Pathologica Microbiologica et Immunologica Scandinavica–Supplement* 1986; **288**:1–151.
7. Larsen RJ, Marx ML. *An Introduction to Mathematical Statistics and its Applications*. Prentice-Hall: Englewood Cliffs, 1986.
8. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics* 1983; **39**:311–324.
9. Haldorsen T, Høie J, Tausjø JE. Cancer. Health region II. Cancer Registry of Norway (in Norwegian), 1990, Oslo.
10. Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine* 1994; **13**:1513–1523.
11. Dyba T, Hakulinen T, Päivärinta L. A simple non-linear model in incidence prediction. *Statistics in Medicine* 1997; **16**:2297–2309.
12. Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H. Trends in cancer incidence and mortality. *IARC Scientific Publications* 1993; **121**:1–806.
13. Osmond C. Using age, period and cohort models to estimate future mortality rates. *International Journal of Epidemiology* 1985; **14**:124–129.

14. Negri E, La Vecchia C, Levi F, Randriamiharisoa A, Decarli A, Boyle P. The application of age, period and cohort models to predict Swiss cancer mortality. *Journal of Cancer Research and Clinical Oncology* 1990; **116**:207–214.
15. Kubik A, Plesko I, Reissigova J. Prediction of lung cancer mortality in four Central European countries, 1990–2009. *Neoplasma* 1998; **45**:60–67.
16. Kiemeney L, Van Berkel H, Witjes JA, Sonke GS, Debruyne FM, Straatman H. Kidney cancer mortality in The Netherlands, 1950–94: prediction of a decreasing trend. *Journal of Epidemiology and Biostatistics* 1999; **4**:303–311.
17. Sharp L, Black RJ, Muir CS, Gemmell I, Finlayson AR, Harkness EF. 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**:1115–1121.
18. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine* 1987; **6**:469–481.
19. Breslow NE, Day NE. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Scientific Publications* 1987; 1–406.
20. Dyba, T. Precision of cancer incidence predictions based on Poisson distributed observations. 2000. Statistical research reports 16. Finnish Statistical Society: Helsinki, 2000.
21. Møller B, Fekjær H, Hakulinen T, Tryggvadóttir L, Storm HH, Talbäck M, Haldorsen T. Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention Supplement* 2002; **11**:S1–S96.
22. Berzuini C, Clayton D. Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* 1994; **13**:823–838.
23. Bray I, Brennan P, Boffetta P. Projections of alcohol- and tobacco-related cancer mortality in Central Europe. *International Journal of Cancer* 2000; **87**:122–128.
24. Teppo L, Hakulinen T, Saxen E. The prediction of cancer incidence in Finland for the year 1980 by means of cancer registry material. *Annals of Clinical Research* 1974; **6**:122–125.
25. Hakulinen T, Teppo L, Saxen E. Do the predictions for cancer incidence come true? Experience from Finland. *Cancer* 1986; **57**:2454–2458.
26. Bashir SA, Esteve J. Projecting cancer incidence and mortality using Bayesian age-period-cohort models. *Journal of Epidemiology and Biostatistics* 2001; **6**:287–296.

