

Projecting the number of patients with colorectal carcinoma by phases of care in the US: 2000–2020

Angela B. Mariotto · K. Robin Yabroff ·
Eric J. Feuer · Roberta De Angelis ·
Martin Brown

Received: 19 December 2005 / Accepted: 6 March 2006
© Springer Science+Business Media B.V. 2006

Abstract

Objective This study provides projections of colorectal cancer prevalence by phases of care (initial, monitoring, and last year of life) to the year 2020 and describes the estimation method.

Methods Cancer prevalence by phase of care was estimated from colorectal cancer incidence and survival from the Surveillance, Epidemiology, and End Results (SEER) Program data, population estimates and projections from the US Census Bureau, and all cause mortality data from the Human Mortality Life Tables. Assumptions of constant incidence and survival were used for projections from 2000 to 2020. Modeled and directly observed patient months by phase of care were compared for 1996–1998 to provide validation of estimates.

Results Prevalence of colorectal cancer is estimated to increase from 1,002,786 (0.36%) patients to 1,522,348 (0.46%) patients between 2000 and 2020. The estimated number of person-months in the initial and last year of life phases of care will increase 43%, while the monitoring phase of care will increase 54%. Modeled person-months by phase of care were consistent with directly observed measures of person months by phase of care in 1996–1998.

Conclusions Under assumptions of current cancer control strategies we project that colorectal cancer prevalence will increase more rapidly than the US

population, largely due to the aging of the US population. This suggests that considerable resources will be needed in the future for initial, continuing and last year of life treatment of colorectal cancer patients unless notable breakthroughs in primary prevention occur in the future years.

Keywords Colorectal cancer · Prevalence · Projection

Introduction

Cancer prevalence, the number of individuals alive ever diagnosed with cancer, is a common measure of the burden of cancer in a population. To better understand the burden of cancer on the health care system, it is important to quantify the population of cancer survivors with respect to phase of care. Prior studies have shown that costs of care are greatest among the newly diagnosed cancer patients during the initial phase of care and among cancer patients in the last year of life phase of care [1]. Because the burden varies across different phases of care, cancer prevalence by phases of care are important for accurately estimating the population treatment and morbidity costs associated with cancer.

The number of cancer survivors has increased steadily since 1970 [2], based on increased incidence and improved survival [3]. These increases are likely to continue based on expected trends of an aging and growing population. As public and research attention is increasingly focused on understanding the implications of current and future cancer control strategies, including risk reduction strategies such as continuous use of aspirin and use of non-steroidal anti-inflammatory

A. B. Mariotto (✉) · K. R. Yabroff · E. J. Feuer ·
R. De Angelis · M. Brown
Division of Cancer Control and Population Sciences,
National Cancer Institute, 6116 Executive Boulevard,
Bethesda, MD 20892, USA
e-mail: mariotta@mail.nih.gov

drugs [4], novel screening technologies such as virtual colonoscopy [5], and innovative treatments [6, 7], projections of cancer prevalence are increasingly important for the purposes of resource allocation, and for planning facilities and manpower requirements for cancer treatment. Several studies have reported current cancer prevalence by phase of care [1, 8] or lifetime costs of colorectal cancer [1, 9], but these approaches only provide prevalence in a single year, and do not allow for projections of cancer prevalence that reflect expected population trends, or change in cancer control strategies. Additionally, these methods often utilize complex claims-based algorithms [1], which may be difficult to develop, validate, and replicate for other tumor sites. The purpose of this study is to describe a method to project cancer prevalence by phase of care and to provide projections of colorectal cancer prevalence to the year 2020. The four main factors that influence prevalence projections are future dynamics of: cancer incidence, cancer survival, age structure and size of population and mortality from other causes. Because it is difficult to anticipate currently undeveloped cancer control technologies and their impact on survival and incidence trends, future projections of incidence and survival are less reliably predicted than increases in population and improvements in other causes mortality. In this paper colorectal cancer prevalence is projected under the assumptions of current levels of incidence and survival and dynamic projections of age and size of the US population and of mortality from other causes. Colorectal cancer care projections under assumption of flat incidence and survival can be interpreted as the effect of the growth and aging of the US population and improvements in other causes of mortality in the colorectal cancer survivors' population, under current cancer control technologies. Thus, these assumptions constitute a starting point for more hypothetical scenarios of future cancer control technologies.

Data and methods

Overview

The approach described here consists of first estimating and projecting total colorectal prevalence. Then cancer prevalence was decomposed into three different phases of care: initial, monitoring, and last year of life. Although in situ cases represent a small proportion of colorectal cancer incidence (approximately 5%), there are costs associated with treatment which consequently burden the health care system. Thus, all the analyses

include both malignant and in situ colorectal tumors. The US population projections from the Census Bureau [10] were used to project incidence rates from the Surveillance, Epidemiology, and End Results (SEER) Program data to the US, and an ecological regression analysis was used to project SEER survival to US. Total colorectal cancer projections through 2020 were estimated using the PIAMOD method [11] that calculates prevalence from cancer incidence and survival. Finally, we validated our estimates of colorectal cancer prevalence in 1996–1998 with direct measures of colorectal cancer incidence from SEER data.

Projecting SEER colorectal incidence cases to the US

We obtained colorectal cancer (malignant and in situ) incidence rates from 1975 to 1999 by single year of age, for males and females from the SEER-9 registries (<http://seer.cancer.gov/>) which include five states (Connecticut, Hawaii, Iowa, New Mexico, Utah) and four metropolitan areas (Atlanta, Detroit, San Francisco-Oakland, and Seattle), representing approximately 10% of the US population. Incidence rates from 2000 to 2020 were assumed to be the same as the 3-year average SEER age specific rate in 1997–1999. The SEER incidence rates were then applied to the respective US population to generate US cancer incidence counts.

US population estimates by single years of age from 1973 to 2022 for males and females were obtained from the US Census Bureau. The middle series (released August, 2002) projections from 1999 to 2022 [10] are based on assumptions about future births, deaths, and international migration. Figure 1 displays male and female populations by 10-year age groups.

Colorectal cancer survival: SEER data and model

Data from 1973 to 1999 from SEER-9 were used to calculate colorectal cancer (malignant and in situ) relative survival rates by period of diagnosis (1973–1977, 1978–1982, 1983–1987, 1988–1991 and 1992–1998), age at diagnosis (15–44, 45–54, 55–64, 65–74 and 75–84) and sex. Cancer survival information is needed for years before 1975 to estimate complete prevalence, i.e. prevalence including all past diagnosis. A parametric cure model [12, 13] was fit to the SEER relative survival and extrapolated beyond the period of observation. This model is described in detail in the appendix. Figures 2A and 2B display the observed and modeled SEER relative survival rates.

Fig. 1 Male and female US populations by age group. US Census Bureau

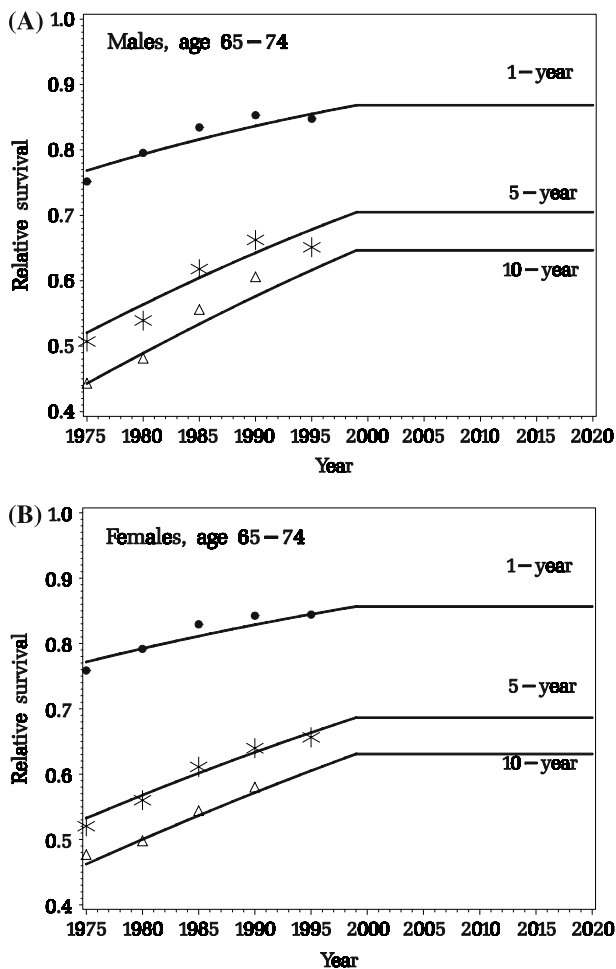
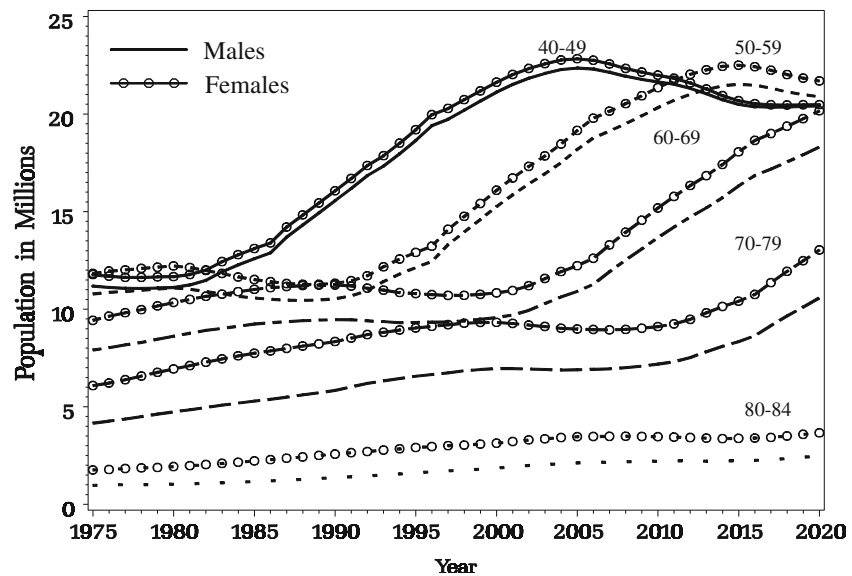


Fig. 2 Colorectal cancer relative survival by follow-up year and age groups. Lines represented modeled survival, assuming constant survival after 1999. Symbols represent observed survival from SEER-9 data

To better reflect colorectal cancer survival in the US, we adjusted the SEER-based colorectal cancer survival by applying a relative risk of 1.06 and 1.05, for males and females, respectively. The relative risks represent 6% and 5% greater risk of dying of colorectal cancer for males and females, respectively, in the entire US compared to SEER areas that were estimated in an ecological analysis on survival [14]. This analysis consisted of regressing SEER survival at the county level on socio-economic variables and extrapolating to the US.

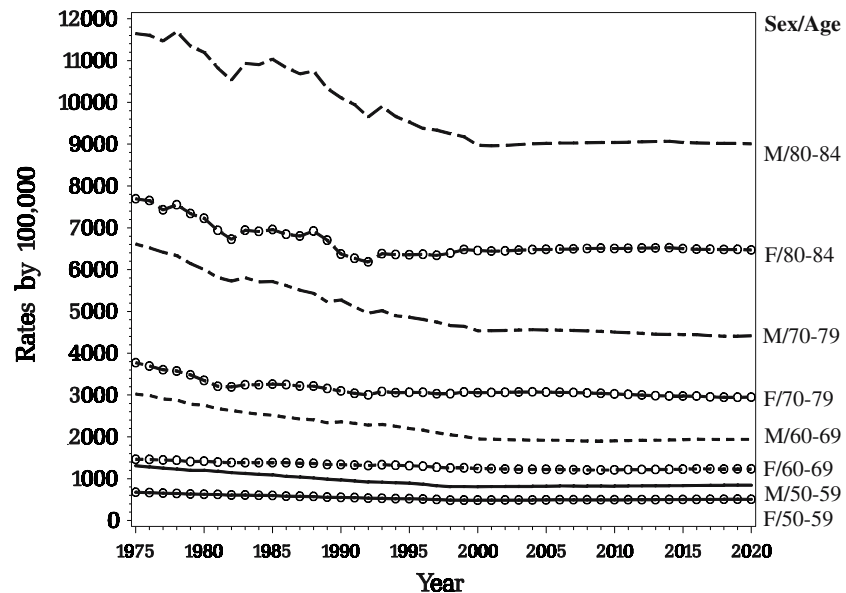
Annual mortality rates from other causes

Annual mortality rates including all causes of death were obtained from the Human Mortality Database [15]. Although these rates include deaths from colorectal cancer, the difference between the all cause mortality rates and rates excluding colorectal cancer deaths are very small and would not affect the analysis here. Mortality from other causes was used to separate the last year of life prevalent cases into those who died of colorectal cancer and those who died of other causes of death. Figure 3 shows trends in other cause mortality rates projected through the year 2020. Age-specific mortality rates for calendar years 2001–2020 were assumed to be the same as those reported for the year 2000.

Prevalence estimation and projections

Prevalence was estimated from US incidence counts and survival model described above using a previously developed method denoted PIAMOD (Prevalence

Fig. 3 All-cause mortality rates (per 100,000) for males and females in the US



Incidence Approach MODEL) [11]. This method is based on an equation that calculates prevalence as the sum of all past cases diagnosed with cancer that survive up to the prevalence date. Let $P_{i,t}$ be the prevalence of those aged i at time t , $I_{j,t-i+j}$ the incidence at age j and time $t - i + j$ and $S(j, t - i + j, i - j)$ the probability of surviving $i - j$ years (to age i) for patients diagnosed at age j and year $t - i + j$. Prevalence is calculated as

$$P_{i,t} = \sum_{j=0}^i I_{j,t-i+j} S(j, t - i + j, i - j). \quad (1)$$

The PIAMOD method is described in detail elsewhere [11]. In order to extrapolate incidence in years before data is available (1975) the PIAMOD method fits an age-period-cohort generalized linear Poisson model to cancer incidence counts. In this study the model was fitted to 1975–2020 incident colorectal cancer cases in the US. A model that best fits the data can be chosen by changing the order of the age, period and cohort polynomials and choosing the one that best fits the data using a statistical test (log-likelihood ratio test). Figure 4 shows the fit of the age, period and cohort model to the observed 1975–2020 colorectal cancer incidence rates for different age groups. Substituting the age-period-cohort incidence model and the survival model previously described, into Eq. (1) PIAMOD calculates prevalence. Prevalence projections were obtained assuming that survival was constant at the same level as estimated in 1999. As mentioned before, although constant incidence and survival may not be realistic assumptions, they represent base assumptions to

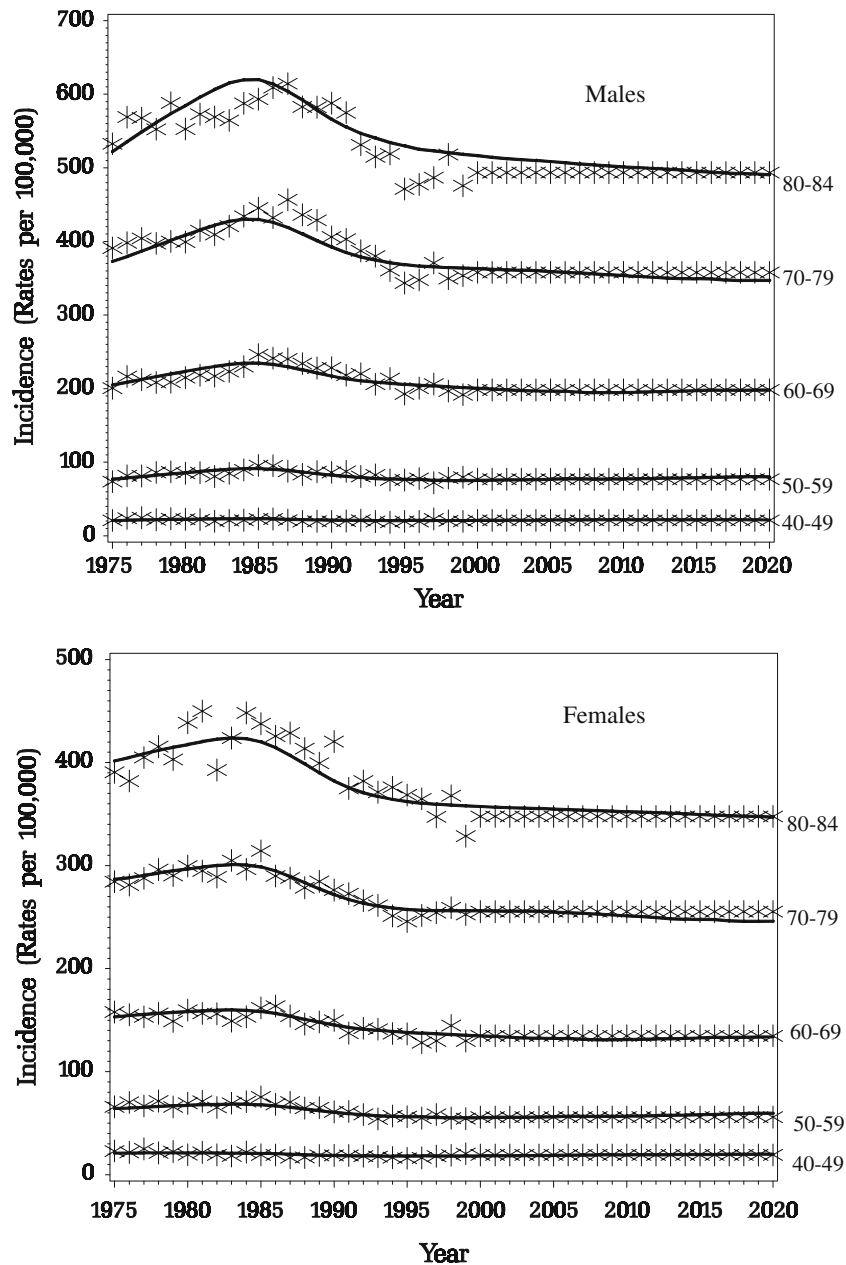
examine the impact of population growth on prevalence projections. Because PIAMOD can only provide results for closed age classes and populations are reported with an older age class of 85+, we produce estimates up to age 84.

Estimating initial, monitoring, and last year of life phase-specific cancer prevalence from total prevalence

Based on prior work estimating the burden of cancer [1], we assigned time after diagnosis to three distinctive phases of care. The initial care phase corresponds to the first 12 months following diagnosis with colorectal cancer, and the last year of life care phase corresponds to 12 months prior to death. The monitoring care phase represents all months between initial and last year of life phases of care. Because the intensity of care for patients with short survival is more similar to patients in the last year of life phase of care [1], we gave the last year of life phase precedence over the other phases of care. For example, if a colorectal cancer patient dies 14 months from diagnosis, the person is in the last year of life phase the 12 months prior to death, in the initial phase the 2 months after diagnosis and never in the monitoring phase.

We first estimated the number of colorectal cancer patients in each phase of care. The number of person-months in each of the phases of care for the years 2000–2020 is calculated by multiplying the estimated number of cases in a particular phase by the estimated average time patients spent in the particular phase during a given year. As a simplifying assumption we assumed that incidence and prevalence are calculated

Fig. 4 Colorectal cancer (including in situ) incidence rates by age groups. Symbols represent SEER rates 1975–1999 and constant assumption after 1999. Lines represent age, period and cohort model



at the beginning of the year (January 1st, year t), and that deaths can occur any time during the year. Estimates provided by PIAMOD (incidence, prevalence, mortality and populations) were calculated at mid year. Estimates at January 1st were calculated by averaging PIAMOD estimates of two consecutive years.

Person-months in initial phase of care

The number of colorectal cancer patients in initial phase in year t is the number of incident cases at year t who survive more than 12 months, because last year of life care has precedence over initial care. A patient

diagnosed at year t dying in the 13th month after diagnosis will be on average 0.5 months in initial care and in the last year of life care the remaining 12 months. If we denote T the time from diagnosis to death and $P[T > t]$ the observed survival distribution for all causes of death for colorectal cancer patients, the average number of months that incident cases at year t are in initial care during year t is calculated as

$$M_1 = \Pr[12 \text{ mth} \leq T < 13 \text{ mth}] \times 0.5 + \Pr[13 \text{ mth} \leq T < 14 \text{ mth}] \times 1.5 + \dots + \Pr[23 \text{ mth} \leq T < 24 \text{ mth}] \times 11.5 + \Pr[T \geq 24 \text{ mth}] \times 12$$

The estimated number of person-months in the initial care phase is given by M_1 multiplied by the incident cases at year t .

Person-months in last year of life phase of care

We distinguished the last year of life phase for patients who died of colorectal cancer and those who died of other causes because of potential differences in costs of care. To estimate the number of person-months among individuals dying of colorectal cancer or other causes of death during the last year of life phase during year t , we first calculated probabilities of death due to colorectal cancer and death due to other causes both in the presence of all causes of death (competing risk framework). Population based survival statistics are usually calculated as net measures of survival, i.e., survival from cancer in the absence of other causes. Because net statistics are not influenced by changes in mortality from other causes these are useful measures for tracking cancer survival progress over time, and making comparisons between racial/ethnic groups or between registries. In this study, we needed to calculate crude probability of death, i.e., the probability of dying of cancer in the presence of other causes of death. This measure reflects mortality patterns actually experienced in a cohort of cancer patients on which many possible causes of death are acting simultaneously. The crude probability of death is reported as a cumulative probability of death rather than survival. Calculation of crude probability of death using life tables has been developed in Cronin and Feuer [16] and can be calculated in SEER*Stat (<http://seer.cancer.gov/seer-stat/>). Calculations for these probabilities are shown in the appendix assuming that mortality for other causes is independent of being diagnosed with cancer.

The number of person-months in last year of life care in year t was calculated by considering incidence at year t and prevalence at year t by years since diagnosis. For each age group we multiply incidence at year t by the probability of dying of CRC and other causes of death within one year. In the same way for each age group we multiplied prevalent cases at January 1st, year t , diagnosed at years $t - 1$, $t - 2$, $t - 3$, $t - 4$ and $t - 5$ to the corresponding probabilities of dying of CRC and other causes within one year given alive at year t and diagnosed at years $t - 1$, $t - 2$, $t - 3$, $t - 4$ and $t - 5$ respectively. For prevalent patients diagnosed more than 5 years before year t , (years $t - 6$, $t - 7$ and so on) we assumed that the hazard of dying of CRC is constant and equal to

the average hazard of people diagnosed 6, 7 and 8 years before. The calculations are shown in detail in the Appendix.

Person-months in monitoring phase of care

The number of person months in the monitoring phase of care during year t is calculated by subtracting from the total person-months prevalence the number of person-months in initial and last year of life care. The total number of person months in care is the number of prevalent patients in year t multiplied by twelve.

Comparisons with observed data from SEER

Direct measures of person-months in the initial, monitoring and last year of life phases of care were obtained from SEER-9 data for the years 1996–1998. For patients diagnosed with a first primary colorectal cancer (invasive or in situ) between 1973 to 1998 with follow-up through 1999, we defined their phase of care for each month in 1996, 1997 and 1998 based on their time of diagnosis and time of death if any, according to the definitions above. For each patient we counted the number of months by calendar year and phase to estimate the number of person months by phase. These more direct estimates of prevalence by phases of care can only be obtained from the year previous to the last year of observed data and do not permit projections. The denominator for the proportion of person-months in each phase of care was the SEER mid-year population multiplied by twelve. These SEER observed proportions were then compared to the proportion of person-months estimated in the US in the years 1996, 1997 and 1998 from the combination of method described above.

Sensitivity analysis

In order to understand the sensitivity of our estimates to the assumptions of constant incidence and survival, we also ran our model under the assumptions that incidence and survival would continue the trend as observed in the last 15 years. Thus we assumed that incidence decreased 1.2%, annually. Using the same notation as in equation [1], incidence at year $(t + 1)$ and age i is calculated as $I_{i,t+1} = I_{i,t} - 0.012I_{i,t}$. We also assumed that survival would improve as estimated by the model described in the Eq. (2) of the Appendix. The term $\exp(-\alpha_2)$ gives the annual relative risk of colorectal death and is estimated to be 0.979 for men and 0.975 for women. Thus, the risk of dying of

colorectal cancer was 2% and 2.5% lower, respectively, for men and women diagnosed with colorectal cancer in a given year compared to people diagnosed a year before.

Results

Table 1 shows that colorectal cancer prevalence is estimated to increase from 1,002,786 (511,743 males and 491,043 females) to 1,522,348 (814,873 males and 707,475 females) between the years 2000 and 2020, under the assumptions of flat incidence, flat survival and dynamic changes in the US population. Additionally, the estimated number of person-months in initial and last year of life care will increase 43%, from 2,200,677 in 2000 to 3,133,704 in 2020, while the number of person-months in monitoring care will increase 54%, from 9,832,746 in 2000 to 15,134,468 in 2020. The monitoring care phase represents the largest phase of care, with approximately 80% to 85% of the total person-months colorectal cancer care prevalence.

Prevalence is higher for males than females, and the difference increases with time reflecting underlying population trends. Figure 5 displays colorectal cancer prevalence by 10-year age groups, which show that the largest increase occurs in the population between ages 70 and 79, in both men and women.

Comparison of the modeled person-month measures of colorectal cancer prevalence by phases of care in the US with equivalent measures over the same time period calculated directly from the SEER-9 data showed very consistent results (Table 2). The percent of person-months in the different phases were almost identical with the exception of last year of life phase for other causes of death in which the US projected estimates were smaller than the SEER observed measures.

In our sensitivity analysis, we used dynamic assumptions of 1.2% decrease in colorectal cancer incidence and 2% (males) and 2.5% (females) annual increase in colorectal cancer survival, and projected 1,431,983 people with colorectal cancer in 2020. This represents 0.44% of the expected US population in 2020, while under the assumption of constant incidence and survival we estimated a prevalence of 0.46% of the US population in 2020.

Discussion

In this study, we adapted an existing method to project colorectal cancer prevalence by phase of care through

the year 2020. Our approach has several advantages over prior work estimating cancer prevalence by phase of care in that our definition of phase of care does not require the use of complex claims-based algorithms [1], and can be used to produce estimates for all colorectal cancer patients of all age groups. Our approach can be used to project colorectal cancer prevalence into the future, rather than just estimate prevalence in current years. The method can be applied to other tumor sites, and combined with other estimates of burden, such as costs of care, to improve the understanding of future cancer burden in the US.

Our findings suggest that the burden of colorectal cancer care may increase dramatically. Under the assumption of current cancer control strategies, projections show that colorectal cancer prevalence will increase 52% from 2000 to 2020 for males and females while the US population is expected to increase approximately 18%. This effect is largely due to the aging of the US population, rather than increases in the size of the US population, because the population over 60 years of age will increase 64% during this period. The highest increase in person-months prevalence is in the monitoring phase (54%), which represents the vast majority of colorectal cancer patients.

Prevalence of colorectal cancer is higher for males than females. We estimated 644,114 male and 569,885 female person-months in initial care phase in 2000, a 13% difference. The colorectal cancer age-adjusted rate is approximately 30% higher among men than women, 59.9 and 45.6 per 100,000 in 2002 [3]. Because women have longer life expectancy than men, life time risk of developing cancer is only 6% higher in men compared to women, 5.84% and 5.51%, respectively [3]. Thus, although men have higher risk of colorectal cancer, the fact that women live longer makes lifetime risk and prevalence measures more similar between men and women.

We were able to validate our modeled estimates against directly calculated measures in 1996–1998. Additionally, our results are largely consistent with those reported elsewhere. The Cancer Statistics Review 1975–2002 [3] reports a total 1,051,682 people ever diagnosed with invasive colorectal cancer as of January 2002 and a total of 873,244 people between ages 0 to 84 (unpublished result). These numbers are direct calculations from the SEER data with an adjustment to estimate the number of people alive who were diagnosed before 1975. The prevalence estimate we obtain is 1,036,294 colorectal cancer patients in 2002 between ages 0 and 84. In this work we include *in situ* colorectal cancers which represent approximately 5% of all CRC reported in SEER.

Table 1 Projections of CRC prevalence and the number and percent of person-months in initial, monitoring and last year of life care phase for the US male and females population aged between 0 and 84

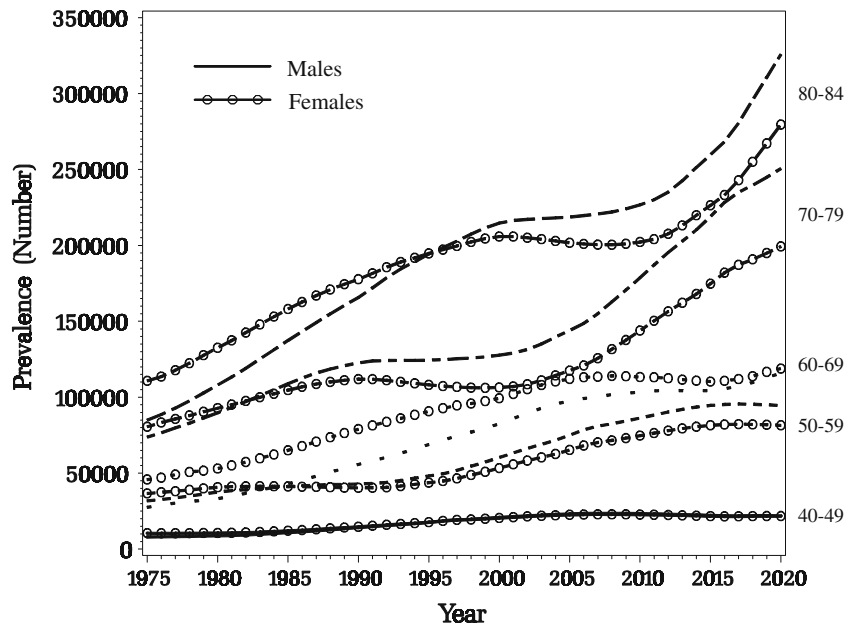
Year	Persons		Person-Months				Percent of person-months			
	Mean population	Prevalence	Initial	Cont.	Last year of life		Initial	Cont.	Last year of life	
					CRC	Other			CRC	Other
<i>Males</i>										
2000	137,171,840	511,743	644,114	5,002,989	252,946	240,863	0.039	0.304	0.015	0.015
2001	138,451,936	521,920	653,685	5,107,615	256,566	245,172	0.039	0.307	0.015	0.015
2002	139,721,792	532,172	663,240	5,213,299	259,991	249,534	0.040	0.311	0.016	0.015
2003	140,992,512	542,628	673,403	5,320,953	263,142	254,039	0.040	0.314	0.016	0.015
2004	142,267,360	553,400	684,318	5,431,605	266,047	258,824	0.040	0.318	0.016	0.015
2005	143,541,056	564,119	695,869	5,541,349	268,457	263,758	0.040	0.322	0.016	0.015
2006	144,811,264	574,906	707,920	5,651,648	270,418	268,882	0.041	0.325	0.016	0.015
2007	146,075,680	586,014	720,586	5,765,156	272,113	274,313	0.041	0.329	0.016	0.016
2008	147,346,720	598,215	734,264	5,889,499	274,549	280,271	0.042	0.333	0.016	0.016
2009	148,612,080	610,832	748,944	6,017,736	276,763	286,541	0.042	0.337	0.016	0.016
2010	149,873,216	624,099	764,377	6,152,564	279,104	293,138	0.043	0.342	0.016	0.016
2011	151,142,848	638,543	780,834	6,299,354	282,097	300,226	0.043	0.347	0.016	0.017
2012	152,419,696	653,820	798,333	6,454,454	285,364	307,684	0.044	0.353	0.016	0.017
2013	153,701,696	669,999	816,694	6,618,625	289,202	315,463	0.044	0.359	0.016	0.017
2014	154,991,376	687,498	836,054	6,796,376	293,819	323,730	0.045	0.365	0.016	0.017
2015	156,276,400	705,717	856,263	6,981,327	298,749	332,270	0.046	0.372	0.016	0.018
2016	157,565,808	725,357	877,292	7,180,867	304,833	341,287	0.046	0.380	0.016	0.018
2017	158,857,376	746,437	899,385	7,394,773	312,371	350,716	0.047	0.388	0.016	0.018
2018	160,146,592	768,681	922,377	7,619,897	321,570	360,333	0.048	0.397	0.017	0.019
2019	161,431,328	791,955	946,048	7,855,466	331,740	370,211	0.049	0.406	0.017	0.019
2020	162,689,952	814,873	969,716	8,087,513	341,217	380,026	0.050	0.414	0.017	0.019
<i>Females</i>										
2000	140,685,488	491,043	569,885	4,829,757	261,119	231,749	0.034	0.286	0.015	0.014
2001	141,921,056	497,456	576,131	4,895,728	262,889	234,726	0.034	0.287	0.015	0.014
2002	143,152,512	504,122	582,517	4,964,532	264,655	237,757	0.034	0.289	0.015	0.014
2003	144,386,608	511,039	589,323	5,035,992	266,267	240,892	0.034	0.291	0.015	0.014
2004	145,620,656	518,047	596,522	5,108,415	267,512	244,110	0.034	0.292	0.015	0.014
2005	146,845,664	524,700	603,891	5,177,207	268,084	247,224	0.034	0.294	0.015	0.014
2006	148,066,512	531,367	611,412	5,246,357	268,245	250,386	0.034	0.295	0.015	0.014
2007	149,280,704	538,202	619,294	5,317,468	267,995	253,671	0.035	0.297	0.015	0.014
2008	150,507,328	546,028	627,919	5,398,637	268,379	257,399	0.035	0.299	0.015	0.014
2009	151,724,976	553,995	637,264	5,481,122	268,330	261,228	0.035	0.301	0.015	0.014
2010	152,938,912	562,408	647,105	5,568,244	268,275	265,273	0.035	0.303	0.015	0.014
2011	154,173,552	572,067	657,874	5,668,026	269,038	269,870	0.036	0.306	0.015	0.015
2012	155,427,008	582,676	669,753	5,777,247	270,212	274,902	0.036	0.310	0.014	0.015
2013	156,694,704	594,205	682,587	5,895,580	271,953	280,338	0.036	0.314	0.014	0.015
2014	157,981,568	607,061	696,495	6,027,374	274,526	286,333	0.037	0.318	0.014	0.015
2015	159,267,712	620,541	711,328	6,165,237	277,344	292,583	0.037	0.323	0.015	0.015
2016	160,571,456	635,525	727,123	6,318,287	281,447	299,439	0.038	0.328	0.015	0.016
2017	161,890,768	652,078	744,209	6,486,726	287,053	306,943	0.038	0.334	0.015	0.016
2018	163,215,296	669,812	762,408	6,666,261	294,150	314,927	0.039	0.340	0.015	0.016
2019	164,545,552	688,787	781,541	6,858,103	302,377	323,417	0.040	0.347	0.015	0.016
2020	165,845,696	707,475	800,991	7,046,955	309,952	331,801	0.040	0.354	0.016	0.017

Because in situ colorectal cancers have a better prognosis than invasive cancers, they represent more than 5% of prevalent colorectal cancer patients, making these estimates comparable.

In a prior study of colorectal cancer prevalence using claims-based algorithms to identify individuals receiving active care, Mariotto [1] estimated that in 1996, 10.4%, 38.1%, 14.1% and 6.3% of colorectal cancer patients received care in the initial, monitoring,

recurrence/metastasis and last year of life phase, respectively. The remaining 37.9% of colorectal cancer patients did not receive any colorectal cancer care in 1996. In the current study, we estimate that approximately 10% and 8% of person-months in a year are in initial and last year of life phase and 82% in the monitoring phase for a similar time period. Our monitoring phase includes people in recurrence, in the monitoring phase, and people not receiving care, and is

Fig. 5 Number of colorectal cancer survivors (1975–1999) by sex and age groups. Projections (2000–2020) based on assumptions of constant survival and incidence after 1999. Males (lines without symbol) Females (lines with symbol)



very consistent with 82.3% of people in recurrence, monitoring and not receiving any colorectal cancer care estimated from the SEER-Medicare data. Because estimates from the SEER-Medicare data are based on more direct observation of phases of care and care received, the similarities confirm robustness of our estimates. Another approach to quantify and qualify cancer prevalence applied to colorectal cancer prevalence in Europe also yields similar findings. Gatta et al. [8] estimated that from the total colorectal cancer patients, 10% were in initial care and 9% in last year of life care in year 1992.

Some of the assumptions used in our study may be overly simplistic. For example, we assumed that incidence occurs at the beginning of each year and that the time in care, irrespective of phase, is 12 months.

However, comparisons with similar studies show that our estimates are consistent with other estimates obtained using different data and methods for the same calendar years. The largest discrepancy observed is when comparing our estimates of the number of person years in last year of life care for other causes of death with directly observed estimates in SEER. Our estimates are smaller than the ones observed in SEER (Table 2). Although the reason for this difference is not clear, the impact of this estimate on the overall cancer burden is small.

Our projections assumed constant colorectal cancer incidence and survival rates at 1999 levels, and US population and other causes mortality projections through 2020. Population and general mortality can be projected up to 2020 with less uncertainty than

Table 2 Number and percent of person-months in initial, monitoring and last year of life care phase for the US and SEER male and females population aged between 0 and 84

Year	Mean population	Percent of person-months				Year	Mean population	Percent of person-months			
		Initial	Cont.	Last year of life				Initial	Cont.	Last year of life	
				CRC	Other					CRC	Other
<i>US modeled estimates—Males</i>						<i>US modeled estimates—Females</i>					
1997	132,373,360	0.039	0.294	0.015	0.014	1997	136,368,096	0.034	0.283	0.016	0.014
1998	133,983,064	0.039	0.297	0.015	0.014	1998	137,838,480	0.034	0.284	0.016	0.014
1999	135,609,568	0.039	0.301	0.015	0.015	1999	139,276,592	0.034	0.285	0.016	0.014
<i>SEER direct calculations—Males</i>						<i>SEER direct calculations—Females</i>					
1996	12,383,218	0.037	0.305	0.014	0.022	1996	12,906,078	0.034	0.316	0.013	0.021
1997	12,505,878	0.038	0.309	0.014	0.023	1997	13,036,528	0.035	0.319	0.013	0.021
1998	12,618,180	0.039	0.315	0.013	0.024	1998	13,153,723	0.037	0.323	0.013	0.023

Comparison between US modeled estimates and direct calculations using the SEER-9 (1975–2000) data

incidence and survival, based on the current population and mortality dynamics. Life styles or economic changes impact populations and mortality more gradually, while changes in cancer control strategies may have more immediate effects. For example, a drop in birth rates in 2000 will only affect the population 50 years and older in 2050. On the other hand, a new effective screening technique that disseminates quickly may change incidence almost immediately. The prevalence estimates presented here, reflect the impact of population dynamics, especially the aging effect and improvements in all causes mortality, on the overall burden of cancer, under currently disseminated cancer control policies and interventions.

As colorectal cancer prevention strategies, such as continuous use of aspirin or non-steroidal anti-inflammatory drugs are adopted by the population, we expect colorectal cancer incidence to decrease. As goals of increasing screening [17] are reached and more efficient technologies become available, we expect a decrease in incidence of invasive disease, from the diagnosis of precancerous adenomas. Increased screening should also lead to a stage shift, with more cancers diagnosed at earlier stages of disease. Because more patients will be diagnosed at earlier stage, survival should also improve. Innovative treatments [7] may also improve survival in the future. Thus, improvements in cancer control strategies may differentially affect prevalence, with decreased incidence leading to *decreased* prevalence, and increased survival leading to *increased* prevalence.

We used sensitivity analysis to explore the effect of assumptions of constant incidence and survival on projections of colorectal cancer prevalence. Applying dynamic assumptions of 1.2% decrease in colorectal cancer incidence and 2% (males) and 2.5% (females) improve in colorectal cancer survival, our results changed little (0.44% vs. 0.46% of the expected US population in 2020). Although survival increased at higher rate than incidence decreased, incidence had a larger impact on prevalence projections.

The method presented in this paper can be used to project cancer prevalence under specific scenarios. However, it can not be used investigate the impact of specific cancer control strategies on cancer survivorship and to estimate the societal return on investments in cancer research. A cooperative agreement funded by the National Cancer Institute, the Cancer Intervention Surveillance Modeling Network (CISNET) (<http://cisnet.cancer.gov/>), uses micro-simulation models to investigate the impact of interventions (i.e. primary prevention, screening, and treatment) on population-based cohorts of patients with breast,

colorectal, prostate and lung cancers. These micro-simulation models require as inputs direct estimates of population use, efficacy, sensitivity and specificity of new interventions, such as screening and treatment. Using the CISNET models, estimates of survival and incidence can be projected using a detailed set of assumptions about the usage patterns of screening and new treatments. They are also able to model directly stage shifts due to dissemination of screening into the population. While these types of projections are undoubtedly more reliable than the projections proposed in this paper, they each involve a large research effort, and therefore can only be done for a very limited number of cancer sites. Integration estimates derived by CISNET models, which can incorporate stage shift, cure and improved survival time into incidence and survival, with the methods of projecting prevalence described here will be an important area for future research.

Prevalence estimates and projections are useful in monitoring the size of the cancer burden in the US, to define care requirements, to establish priorities, and to project costs of cancer care. In this paper we project that the population diagnosed with colorectal cancer will increase more rapidly than the overall US population, under assumptions of current cancer control strategies, as well as under assumptions of continuing trends in incidence and survival as observed in the most recent years. This suggests that considerable resources will be needed in the future for initial, continuing and last year of life treatment of colorectal cancer patients unless notable breakthroughs in primary prevention occur in the future years.

Appendix

Survival model used in PIAMOD to estimate prevalence

The survival model used in PIAMOD assumes a proportion $\phi(i)$ of patients, dependent on age at diagnosis i , who are cured from the cancer and die with the same mortality rate as the general population. The remaining have a relative survival following a Weibull distribution with parameters $\beta(i)$ and $\lambda(i)$. Mathematically, the model specifies that for patients in the birth cohort $t - i$, the probability of surviving age $i + d$ for patients diagnosed at age i and year t is given by,

$$S(i, t, d) = \left\{ \phi(i) + [1 - \phi(i)] \exp - [\lambda(i)d]^{\beta(i)} \right\}^{\exp[-\alpha_2(t-t_0) - \alpha_3]}, \quad (2)$$

where $\exp[-\alpha_2]$ is the annual relative risk of cancer death for patients diagnosed in calendar year $t + 1$ with respect to period t and $\exp[-\alpha_3]$ a relative risk of death of US patients with respect to SEER patients. The parameters $\phi(i)$, $\lambda(i)$, $\beta(i)$ and α_2 are estimated in CANSURV (<http://srab.cancer.gov/cansurv/>) a statistical software to analyze population-based survival data. Parameters α_3 were estimated by using an ecological regression model to the SEER relative rates and extrapolating to the US [14]. More specifically, the association of demographic-socio-economic (ecological) variables at county level on cancer relative survival rates are estimated in SEER and then extrapolated to the US. For all races, the relative risks of cancer death in US with respect to SEER were 1.12 for breast, 1.15 for prostate, 1.06 for colorectal males and 1.05 for colorectal female cancers. Figure 3 displays the observed and modeled relative survival rates some combination of age at diagnosis, period of diagnosis and time from diagnosis. Survival is assumed constant after 1999.

Calculation of the probabilities of death in the presence of all causes of death

For people diagnosed with colorectal cancer at age j , we define h_{ji} the net (in the absence of other causes) probability of dying of cancer between ages i and $i + 1$, conditioned that they are alive prior to age i . Since we are assuming a constant colorectal cancer survival after 1999 we do not need to consider an index of time. This probability is calculated from s_{ji} the interval relative survival estimated from SEER-9 cases diagnosed in 1991–1999.

$$h_{ji} = (1 - s_{ji})$$

Similarly, the probability of dying from other causes, between ages i and $i + 1$ at year t , conditioned they survived prior to age i , \bar{h}_i^t , can be calculated as

$$\bar{h}_i^t = [1 - E_i^t]$$

where E_i^t is the US life tables from the Human Mortality Database [15]. In this calculation we assume that mortality from other causes is independent of having cancer. In other words, a person diagnosed with cancer has the same hazard of dying of other causes as a person free of cancer. This assumption might not be true if we believe the disease or its treatment might affect other causes mortality. Age and year of birth are the important factors affecting mortality for other causes. The crude (in the presence of all causes of death) probabilities of dying for colorectal cancer and other causes of death between ages

i and $i + 1$ at year t , given they are alive just prior to age i and were diagnosed with cancer at age j , denoted $\delta_{i,t}^j$ and $\varepsilon_{i,t}^j$, respectively, are calculated as

$$\delta_{i,t}^j = h_{ji} - \frac{1}{2}h_{ji} \bar{h}_i^t \text{ and } \varepsilon_{i,t}^j = \bar{h}_i^t - \frac{1}{2}h_{ji} \bar{h}_i^t.$$

Details of calculations can be found in Cronin and Feuer [16].

Details of calculation to estimate number of person months in last year of life phase of care

We divide prevalent cases alive at January 1st year t by years since diagnosis. We first consider incidence cases diagnosed at January 1st year t , $I_{i,t}$ and apply the probabilities of dying in the same year $\delta_{i,t}^i$ and $\varepsilon_{i,t}^i$. We then consider prevalence by years since diagnosis. For prevalent cases diagnosed $d = 1, \dots, 5$ years prior to the prevalence date (diagnosed at year $t - d$ or age $i - d$ and denoted $P_{i,t}^{t-d}$), the quantities $P_{i,t}^{t-d} \delta_{i,t}^{i-d}$ and $P_{i,t}^{t-d} \varepsilon_{i,t}^{i-d}$ represent the estimated number dying during year t due to cancer and other causes, respectively, and $\sum_{d=1}^5 P_{i,t}^{t-d} \delta_{i,t}^{i-d}$ and $\sum_{d=1}^5 P_{i,t}^{t-d} \varepsilon_{i,t}^{i-d}$, represent the estimated number dying during year t for individuals diagnosed in the previous 5 years. Prevalent cases diagnosed more than 5 years from the prevalence date can be calculated as $P_i - \sum_{d=1}^5 P_{i,t}^{t-d}$. Assuming that after 5 years from diagnosis the annual hazard of dying of colorectal cancer is constant we apply and average hazard estimated using the hazards at year 6, 7 and 8 from diagnosis to estimate the number dying among those diagnosed 5 years or more from the prevalence date

$$\left\{ P_i - \sum_{d=1}^5 P_{i,t}^{t-d} \right\} \frac{1}{3} \sum_{d=6}^8 \delta_{i,t}^{i-d}.$$

Thus, the number of CRC deaths at that year t is calculated as

$$I_{i,t} \delta_{i,t}^i + \sum_{d=1}^5 P_{i,t}^{t-d} \delta_{i,t}^{i-d} + \left\{ P_i - \sum_{d=1}^5 P_{i,t}^{t-d} \right\} \frac{1}{3} \sum_{d=6}^8 \delta_{i,t}^{i-d}$$

and the number of deaths due to causes other than colorectal at year y is

$$I_{i,t} \varepsilon_{i,t}^i + \sum_{d=1}^5 P_{i,t}^{t-d} \varepsilon_{i,t}^{i-d} + \left\{ P_i - \sum_{d=1}^5 P_{i,t}^{t-d} \right\} \frac{1}{3} \sum_{d=6}^8 \varepsilon_{i,t}^{i-d}.$$

As a simplifying assumption, we consider that these patients contribute 12 months to the terminal care phase in year t because they represent a snapshot of the histories in the continuous timeline.

References

1. Brown ML, Riley GF, Potosky AL, Etzioni RD (1999) Obtaining long-term disease specific costs of care: application to medicare enrollees diagnosed with colorectal cancer. *Med Care* 37(12):1249–1259
2. CDC. Cancer Survivorship – United States (1971–2001) *MMWR* 6-25-2004; 53(24):526–529
3. SEER Cancer Statistics Review (1975–2002) Ries LAG, Eisner MP, Kosary CL, et al. (2005) National Cancer Institute (http://seer.cancer.gov/csr/1975_2002).
4. Garcia-Rodriguez LA, Huerta-Alvarez C (2001) Reduced risk of colorectal cancer among long-term users of aspirin and nonaspirin nonsteroidal antiinflammatory drugs. *Epidemiology* 12(1):88–93
5. Pickhardt PJ, Choi JR, Hwang I, et al. (2003) Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 349(23):2191–2200
6. Cunningham D, Humblet Y, Siena S, et al. (2004) Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* 351(4):337–345
7. Poynter JN, Gruber SB, Higgins PD, et al. (2005) Statins and the risk of colorectal cancer. *N Engl J Med* 352(21):2184–2192
8. Gatta G, Capocaccia R, Berrino F, Ruzza MR, Contiero P (2004) The EUROPREVAL working group. Colon cancer prevalence and estimation of differing care needs of colon cancer patients. *Ann Oncol* 15(7):1136–1142
9. Mariotto A, Warren JL, Knopf KB, Feuer EJ (2003) The prevalence of patients with colorectal carcinoma under care in the US. *Cancer* 98(6):1253–1261
10. U.S. Census Bureau, Population Division. Interim projections consistent with Census 2000 (released March 2002) <http://www.census.gov/population/www/projections/pop-proj.html.2002>
11. Verdecchia A, De Angelis G, Capocaccia R (2002) Estimation and projections of cancer prevalence from cancer registry data. *Stat Med* 21(22):3511–3526
12. Goldman AI (1984) Survivorship analysis when cure is a possibility: a Monte Carlo study. *Stat Med* 3(2):153–163
13. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A (1999) Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med* 18(4):441–454
14. Mariotto A, Capocaccia R, Verdecchia A, et al. (2002) Projecting SEER cancer survival rates to the US: an ecological regression approach. *Cancer Causes Control* 13(2):101–111
15. Human Mortality Database. University of Californian, Berkeley (USA), and Mark Planck Institute for Demographic Research (Germany). Available at <http://www.mortality.org> (data downloaded on August 2002)
16. Cronin KA, Feuer EJ (2000) Cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Stat Med* 19:1729–1740
17. Winawer S, Fletcher R, Rex D, et al. (2003) Colorectal cancer screening and surveillance: clinical guidelines and rationale-update based on new evidence. *Gastroenterology* 124(2):544–560