STATISTICS IN MEDICINE
*Statist. Med.* 2000; **19**:1729–1740

# Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival‡

Kathleen A. Cronin*,† and Eric J. Feuer

*Applied Research Branch, National Cancer Institute, EPN 4103, 6130 Executive Boulevard, Bethesda, MD 20892, U.S.A.*

SUMMARY

A common population-based cancer progress measure for net survival (survival in the absence of other causes) of cancer patients is relative survival. Relative survival is defined as the ratio of a population of observed survivors in a cohort of cancer patients to the proportion of expected survivors in a comparable set of cancer-free individuals in the general public, thus giving a measure of excess mortality due to cancer. Relative survival was originally designed to address the question of whether or not there is evidence that patients have been cured. It has proven to be a useful survival measure in several areas, including the evaluation of cancer control efforts and the application of cure models. However, it is not representative of the actual survival patterns observed in a cohort of cancer patients. This paper suggests a measure for cumulative crude (in the presence of other causes) cause-specific probability of death for a population diagnosed with cancer. The measure does not use cause of death information which can be unreliable for population cancer registries. Point estimates and variances are derived for crude cause-specific probability of death using relative survival instead of cause of death information. Examples are given for men diagnosed with localized prostate cancer over the age of 70 and women diagnosed with regional breast cancer using Surveillance, Epidemiology and End Results (SEER) Program data. The examples emphasize the differences in crude and net mortality measures and suggest areas where a crude measure is more informative. Estimates of this type are especially important for older patients as new screening modalities detect cancers earlier and choice of treatment or even 'watchful waiting' become viable options. Published in 2000 by John Wiley & Sons, Ltd.

## 1. INTRODUCTION

Many population-based cancer statistics have both a net (in the absence of other causes) and a crude (in the presence of other causes) formulation. For population-based cancer survival, relative survival is used as a net measure. Surprisingly there is no crude population-based survival measure currently reported. In this paper we develop a new population-based cancer survival statistic,

*Correspondence to: Kathleen A. Cronin, Applied Research Branch, National Cancer Institute, EPN 4103, 6130 Executive Boulevard, Bethesda, MD 20892, U.S.A.
† E-mail: cronink@dcpcepn.nci.nih.gov


*Received November 1998*
Published in 2000 by John Wiley & Sons, Ltd. *Accepted August 1999*

analogous to relative survival, that measures mortality in the presence of other causes without the use of cause of death information.

In a population with multiple causes of death acting simultaneously, cumulative mortality may be represented in the absence or presence of other causes based on the theory of competing risks [1, 2]. Historically, motivation for the development of methods to estimate survival in the absence of other causes was closely related to the idea of 'cure' [3]. Researchers were interested in studying if and when expected survival for cancer patients returned to the level of the general population. More recently, researchers have been interested in using population-based estimates in the absence of other causes in policy decisions, such as tracking the progress of cancer control efforts, since it is not influenced by changes in mortality from other causes [4]. For example, progress in heart disease should not obscure recent advances in breast cancer treatment. Because of its desirable properties (that is, interpretation as excess mortality due to cancer and unaffected by changes in mortality of other diseases), survival in the absence of other causes is often the only population based cancer measure reported representing patient survival. Therefore, it is interpreted as the mortality impact of a cancer diagnosis, which is not consistent with the original intent of the measure. In the absence of other causes, survival estimates are not representative of the mortality patterns actually experienced in a cohort of cancer patients.

Gaynor *et al.* [5] contrast the use of cause-specific cumulative mortality in the absence and presence of other causes of death where cause of death is known. To date, little attention has been paid to the derivation of a population-based estimate of cumulative mortality in the presence of other causes when cause of death information is unreliable or unknown. In Section 2 of this paper we review the theory of competing causes as it applies to cancer survival. In Section 3 we derive estimates of crude cause-specific probability of death without using cause of death information and discuss potential problems with this measure when it is estimated across heterogeneous strata or when the population lifetable does not match the cancer population. Section 4 provides two examples contrasting net and crude cumulative mortality.

## 2. COMPETING CAUSES OF DEATH AND CANCER SURVIVAL

Various terminology has appeared in the literature for survival and mortality in the absence and in the presence of other causes. Throughout this paper we use the terms net survival and net probability of death in the absence of other causes, and crude survival and crude probability of death in the presence of other causes. Survival estimates in the absence of other causes, that is, the net survival function, can be represented by $S_k(r) = \exp^{-\int_0^r \lambda_k(t)\,\mathrm{d}t}$, where $\lambda_k(t)$ is the net hazard rate for cause $k$. Cumulative mortality for cause $k$ is estimated as $H_k(r) = 1 - S_k(r)$. The calculation of $H_k(r)$ assumes that only one cause of death is acting in the population, thus the net hazard rate $\lambda_k$ equals the total hazard rate for the population.

Crude cumulative mortality is defined as $G_k(r) = \int_0^r S(t)v_k(t)\,\mathrm{d}t$ [1], where $S(t)$ is survival from all causes until time $t$, $v_k(t)$ is the crude hazard rate (in the presence of other causes) for cause $k$, and $S(t)v_k(t)$ represents surviving all causes up until time $t$ and subsequently dying of cause $k$ at time $t$. At any time $r$ the population can be classified into one of four mutually exclusive and exhaustive groups: die of cause $k$ at or before time $r$; die of other causes at or before time $r$; die of cause $k$ after time $r$, and die of other causes after time $r$. Mathematically this can be

expressed as

$$\int_0^r S(t)v_k(t)\,dt + \sum_{i\neq k}\int_0^r S(t)v_i(t)\,dt + \int_r^\infty S(t)v_k(t)\,dt + \sum_{i\neq k}\int_r^\infty S(t)v_i(t)\,dt = 1$$

where $i = 1 \ldots K$ are competing causes of death. Crude cause-specific survival, $(1 - G_k(r)) = \sum_{i\neq k}\int_0^r S(t)v_i(t)\,dt + \sum_{i=1}^K \int_r^\infty S(t)v_i(t)\,dt$, is the probability of dying of any cause other than $k$ prior to time $r$ or dying of any cause (including cause $k$) after time $r$. Because of this awkward interpretation of crude survival, we estimate cumulative mortality rather than survival when working in the presence of other causes.

The sum of net mortality from all possible causes of death does not equal observed mortality, that is, $\sum_k(1 - S_k(r)) \neq (1 - S(r))$. However, crude cumulative probability of death has the intuitive property of partitioning the total observed mortality by cause of death. For example, if $G(r)$ represents total observed mortality, $G(r) = 1 - S(r)$, then $\sum_k G_k(r) = G(r)$. When independence of competing risks in the population is assumed, the net hazard is equal to the crude hazard for both cancer and other causes ($\lambda_k(t) = v_k(t)$).

Population-based cancer registries do not generally have continuous time follow-up that would allow for the direct application of the survival models described above. Population-based cancer survival statistics are collected in discrete time interval with many events (deaths and censors) in a single time interval. Therefore, interval-based methods such as lifetables are used to report results. Another consideration when working with population-based cancer registries is that cause of death is ascertained from death certificates which are sometimes missing, have an unknown cause of death recorded, or are not sufficiently reliable for the estimation of cause-specific survival [6, 7]. Without cause of death information, survival from cancer in the absence of other causes is estimated using relative survival [8].

Relative survival is based on the assumption that competing risks in the population are acting independently, so that $S(t) = S_c(t)S_o(t)$ where $S(t)$ is the observed survival from all causes in a cohort of cancer patients, $S_c(t)$ is net cause-specific survival for cancer and $S_o(t)$ is the net survival for non-cancer causes. Relative survival estimates net survival from cancer using the ratio of the proportion of observed survivors in a cohort of cancer patients to the proportion of expected survivors in a comparable set of cancer-free individuals in the general population, that is, $\frac{S(t)}{S_o(t)}$.

In the following section we apply the theory of competing risks to the situation encountered when using data from population-based cancer registries. To be consistent with the available data, all terms are defined for interval data. We estimate the conditional probability of survival in a specified interval given that an individual was alive at the beginning of the interval, $\exp^{-\int_x^{x+1}\lambda_k(t)dt}$, rather than working directly with the instantaneous hazard functions $\lambda_k(t)$.

## 3. ESTIMATING CRUDE PROBABILITY OF DEATH

For a discrete time interval $x$, we calculate the crude probability of death due to cancer and other causes ($\tilde{g}_{xc}$ and $\tilde{g}_{xo}$) and the crude cumulative cause-specific probability of death ($\tilde{G}_{xc}$ and $\tilde{G}_{xo}$). We use the term 'probability of death' since we are discussing case mortality rather than population mortality, and the notation of $\tilde{g}$ and $\tilde{G}$ (rather than the usual '^' notation) to emphasize that the estimates are obtained without the use of cause of death information.

Assume that survival data are grouped by interval of follow-up through time $M$ after diagnosis. We consider $m$ time intervals from diagnosis given by $I_x = [a_x, a_{x+1})$ for $x = 1 \ldots m$, $a_1 = 0$ and $a_{m+1} = M$. Data may be stratified by a number of covariates such as age, race, sex and year of diagnosis. The issue of obtaining pooled estimates across heterogeneous covariate classes is discussed in detail in Section 3.2. The following data are usually available as output from a relative survival computer package [9, 10]:

$n_x =$ number alive at the beginning of interval $x$

$d_x =$ number died in interval $x$

$l_x =$ number lost to follow-up in interval $x$

$E_x =$ expected net survival for other causes in interval $x$ conditioned on being alive at the beginning of interval $x$

The number of people at risk during the interval is adjusted for uniform loss to follow-up, $n_x^* = n_x - \frac{1}{2} l_x$. The probability of surviving interval $x$ conditioned on surviving until the beginning of the interval is estimated using a lifetable approach [11] and assumed to be a binomial random variable with a maximum likelihood estimator of $\hat{P}_x = (1 - \frac{d_x}{n_x^*})$.

$E_x$ is the conditional expected survival for interval $x$ for a set of individuals from the general population comparable to the cohort of cancer patients. Thus $E_x$ estimates the net survival from other causes for the cohort of cancer patients (that is, the survival that the cohort would have expected if they did not have the cancer under study). Let $E_x^o(\psi)$ be the cumulative expected survival for $x$ intervals from the general population matched to a cohort of cancer patients on age in the year of diagnosis, sex and race, $\psi =$ (age at diagnosis, sex, race, year of diagnosis). The expected cumulative survival for the cohort of cancer patients is defined as $E_x^o = \Sigma_\psi w_\psi E_x^o(\psi)$, where $w_\psi$ is the probability of being diagnosed with the covariate profile $\psi$ in the cohort of cancer patients [8]. Conditional expected survival $E_x = E_x^o / E_{x-1}^o$. $E_x$ is considered to be fixed rather than a random variable.

The estimates derived in this paper assume independent competing causes of death. For mathematical convenience, we use the concept of a latent time of death for each competing cause acting within a population. A latent time for cause $k$ is defined as the time death would occur from cause $k$ in the absence of all other causes of death. The probability of dying in interval $x$ conditioned on surviving until the beginning of the interval can be written as $1 - S_{xc} S_{xo}$, where $S_{xc} S_{xo}$ is the probability of surviving both cancer and other causes. This can also be written as $1 - (1 - h_{xc})(1 - h_{xo}) = h_{xc} + h_{xo} - h_{xc} h_{xo}$, where $h_{xc}$ and $h_{xo}$ are the probabilities that the latent time of death (that is, net probabilities of dying) for cancer and other causes occurs in interval $x$, respectively. The last term represents the probability that the latent time of death due to cancer and the latent time of death due to other causes fall within the same interval.

When estimating the crude probability of death due to cancer or other causes, a problem arises as to how to divide the probability that both latent times of death fall in the same interval ($h_{xc} h_{xo}$) between cancer and other causes. Hakulinen [12] reviews a number of approaches to modelling the relationship between crude and net survival. We use the ratio model formulation described in Hakulinen's paper. The ratio model gives the probability of dying of cancer as $h_{xc} - r h_{xc} h_{xo}$, where $r$ is the proportion of the probability $h_{xc} h_{xo}$ that is assigned to cancer. Correspondingly $(1 - r)$ is the proportion of $h_{xc} h_{xo}$ that is assigned to other causes. We assume an approximate uniform distribution for the time of death from cancer and other causes over each interval, $h_{xc} \sim$ uniform $[a_x, a_{x+1})$ and $h_{xo} \sim$ uniform $[a_x, a_{x+1})$. In this case $P(h_{xc} < h_{xo} | h_{xc}$ and $h_{xo} \in$ interval $x) = \frac{1}{2}$, so half the time the cancer death will occur before death from other causes when both occur in the same

interval. Therefore, the chance of dying of cancer in interval $x$ conditioned on surviving until the beginning of the interval is $h_{xc} - \frac{1}{2}h_{xc}h_{xo} = h_{xc}(1 - \frac{1}{2}h_{xo})$ and the chance of dying of other causes conditioned on surviving until the beginning of the interval is $h_{xo} - \frac{1}{2}h_{xc}h_{xo} = h_{xo}(1 - \frac{1}{2}h_{xc})$.

We set $h_{xo} = (1 - E_x)$ and $h_{xc} = (1 - \frac{\hat{P}_x}{E_x})$, where $E_x$ is based on U.S. lifetables and relative survival ($\frac{\hat{P}_x}{E_x}$) is an estimate of net survival from cancer. Ideally deaths from the type of cancer under study would be identified and treated as censored when estimating net survival from other causes. Although U.S. lifetables include cancer deaths, the number of deaths due to any one type of cancer is small and $E_x$ is a reasonable approximation for the probability of death from causes other than the cancer under study. Thus the interval formulation of cause-specific probabilities of death from cancer and other causes for cancer patients are

$$\tilde{g}_{xc} = \left(\prod_{i=1}^{x-1} \hat{P}_i\right)\left(1 - \frac{\hat{P}_x}{E_x}\right)\left(1 - \frac{1}{2}(1 - E_x)\right)$$

$$\tilde{g}_{xo} = \left(\prod_{i=1}^{x-1} \hat{P}_i\right)(1 - E_x)\left(1 - \frac{1}{2}\left(1 - \frac{\hat{P}_x}{E_x}\right)\right)$$

The variances of $\tilde{g}_{xc}$ and $\tilde{g}_{xo}$ are found by using the delta method and then replacing $P_x$ by its MLE estimate. The variance equations, which is an analogue of Greenwood's formula, are

$$\widehat{\text{var}}(\tilde{g}_{xc}) = \tilde{g}_{xc}^2\left(\sum_{i=1}^{x-1}\left(\frac{d_i}{(n_i^* - d_i)n_i^*}\right) + \left(\frac{\hat{P}_x}{E_x - \hat{P}_x}\right)^2\left(\frac{d_x}{(n_x^* - d_x)n_x^*}\right)\right)$$

$$\widehat{\text{var}}(\tilde{g}_{xo}) = \tilde{g}_{xo}^2\left(\sum_{i=1}^{x-1}\left(\frac{d_i}{(n_i^* - d_i)n_i^*}\right) + \left(\frac{\hat{P}_x}{E_x + \hat{P}_x}\right)^2\left(\frac{d_x}{(n_x^* - d_x)n_x^*}\right)\right)$$

The cumulative estimates for cause-specific probability of death due to cancer and other causes ($\tilde{G}_{xc}$ and $\tilde{G}_{xo}$), are defined as $\sum_{i=1}^{x}\tilde{g}_{ic}$ and $\sum_{i=1}^{x}\tilde{g}_{io}$, respectively. Since the probability of dying from cancer in any one interval is conditioned on surviving previous intervals, $\tilde{g}_{ic}$ and $\tilde{g}_{jc}$ are dependent random variables. The variance of the cumulative estimate is $\sum_{i=1}^{x}\text{var}(\tilde{g}_{ic}) + 2\sum_{k=1}^{x}\sum_{j=k+1}^{x}\text{cov}(\tilde{g}_{kc}, \tilde{g}_{jc})$. In estimating the covariance terms we ignore all terms of order $n_x^{*-2}$ and again replace $P_x$ with $\hat{P}_x$. The covariance term is shown below for $k < j$:

$$\widehat{\text{cov}}(\tilde{g}_{kc}, \tilde{g}_{jc}) = \left(\prod_{i=1}^{k-1}\hat{P}_i^2\right)\left(\prod_{i=k}^{j-1}\hat{P}_i\right)(1 - \tfrac{1}{2}(1 - E_k))(1 - \tfrac{1}{2}(1 - E_j))\left(1 - \frac{\hat{P}_k}{E_k}\right)\left(1 - \frac{\hat{P}_j}{E_j}\right)$$
$$\times\left(-\frac{(1 - \hat{P}_k)}{(E_k - \hat{P}_k)n_k^*} + \sum_{i=1}^{k-1}\frac{d_i}{(n_i^* - d_i)n_i^*}\right)$$

Similarly, the variance term for the $\tilde{G}_{xo}$ is $\sum_{i=1}^{x}\text{var}(\tilde{g}_{io}) + 2\sum_{k=1}^{x}\sum_{j=k+1}^{x}\text{cov}(\tilde{g}_{ko}, \tilde{g}_{jo})$ with the covariance term for $k < j$ shown below:

$$\widehat{\text{cov}}(\tilde{g}_{ko}, \tilde{g}_{jo}) = \left(\prod_{i=1}^{k-1}\hat{P}_i^2\right)\left(\prod_{i=k}^{j-1}\hat{P}_i\right)(1 - E_k)(1 - E_j)\left(1 - \frac{1}{2}\left(1 - \frac{\hat{P}_k}{E_k}\right)\right)\left(1 - \frac{1}{2}\left(1 - \frac{\hat{P}_j}{E_j}\right)\right)$$
$$\times\left(\frac{(1 - \hat{P}_k)}{(E_k + \hat{P}_k)n_k^*} + \sum_{i=1}^{k-1}\frac{d_i}{(n_i^* - d_i)n_i^*}\right)$$

### 3.1. Heterogeneity in the cohort under study

Issues of estimating relative survival across heterogeneous strata have been discussed by Hakulinen [13, 14], Hakulinen and Abeywickrama [10] and Esteve *et al.* [15]. It is known [14] that when survival for both cancer and other causes are different across strata, the estimate of population relative survival is biased and tends towards the relative survival of the strata with the longest survival.

The Appendix gives the procedure for pooling strata specific estimates and describes when pooled estimates are required. The results show that ignoring strata when estimating $g_{xc}$ is biased only when both $h_{xc}$ and $h_{xo}$ vary across strata. Survival from other causes will differ by age strata, especially if older age groups are included. However, for many cancers, survival from cancer may not differ significantly by age, and in these cases an unbiased estimate can be obtained without pooling separate age specific strata.

### 3.2. Match of the cancer population to the general population and estimates of $g_{xc}$ less than zero

It is not unusual to observe relative survival greater than 1, which forces $g_{xc}$ to be less than zero. Negative estimates for $g_{xc}$ occur when the expected number of deaths in an interval $x$ is greater than the observed number. This can occur in cases where the number at risk is small, due to high variance of the estimate based on small numbers and the fact that observed deaths must occur in whole increments while expected deaths derived from population lifetables can occur in any fraction. It may also be that the general population lifetable is not properly matched to the observed cancer population. Cancer patients may have lower or higher mortality from other causes than the general population. A comparison between relative survival and cause-specific survival has been previously performed using several different definitions of a cancer death to account for the inaccuracies associated with death certificate information [16]. The relative and cause-specific survival estimates generally give similar results, indicating that general population expected survival does approximate expected survival for a cancer cohort. Exceptions to this occur when there is one dominant risk factor that results in mortality from multiple diseases, such a lung cancer where smoking is the dominant risk factor.

Similar to the adjustment made to relative survival, it is possible to make an adjustment for $g_{xc}$ less than zero to ensure that $G_{xc}$, is non-decreasing; $\tilde{g}_{xc}$ could be set equal to zero for the interval where the estimate is negative and for calculating cumulative estimates in subsequent intervals. The variance of $\tilde{g}_{xc}$ and the covariance terms which include that interval would also be set to zero. An analogous adjustment would be made for the estimates of $\tilde{g}_{xo}$ and associated variances assigning all deaths in the interval to other causes. Although this procedure has the advantage of guaranteeing positive probability estimates, the adjustment can lead to bias in estimating cumulative probability of death. When $g_{xc}$ is actually zero the estimate should fall around zero; adjusting the negative errors to zero and leaving the positive errors alone will bias the cumulative estimate.

## 4. EXAMPLES

Survival information following cancer diagnosis was obtained from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER)Program [9]. SEER is a geographically defined, population-based tumour registry comprising 13.9 per cent of the U.S. population. Cases

were diagnosed from 1973 to 1994 and followed to 1994. Computations were done using data output from the SEER*STAT Survival System developed at the National Cancer Institute, Division of Cancer Control and Population Sciences, Cancer Statistics Branch. Crude cause-specific probability of death calculations will be publicly available in upcoming versions of SEER*STAT(see web site http://www-seer.ims.nci.nih.gov/ScientificSystems/SEERStat/ for more information). Individuals were classified using SEER historical summary stage of disease at diagnosis. Localized is defined as disease confined entirely to the organ of origin, and regional is disease that has extended beyond the limits of the organ of origin into surrounding organs or tissue and/or regional lymph nodes. We look at two examples using SEER data that highlight the differences between net and crude cause-specific probability of death. The examples chosen demonstrate the impact of competing causes of death in older age groups on the probability that a cancer patient dies from their cancer.

### *4.1. Localized prostate cancer in men over the age of 70*

Table I demonstrates the calculations for men 70 years and older diagnosed with localized prostate cancer between 1973 and 1994 with intervals defined as years. The first eight columns are output generally available from cancer registries. The information from the first eight columns along with the equations derived in this paper are used to calculate point estimates and variances for crude cause-specific probability of death from cancer and other causes, shown in the last ten columns of the table. Note that $1 - \prod_{i=1}^{x} P_i = G_{xc} + G_{xo}$ produces a partitioning of the overall cumulative mortality. Ninety per cent of these men die within 15 years of diagnosis, 18 per cent of prostate cancer and 72 per cent of other causes. The standard errors for $G_{xc}$ and $G_{xo}$ become larger toward the end of follow-up where the number at risk becomes small. Figure 1 graphs the information in Table I to compare net and crude cumulative mortality for these patients. The graph on the left shows net cumulative cancer specific probability of death (1-relative survival) over a fifteen year period. The graph on the right of Figure 1 plots observed mortality for all causes and an estimate of crude cumulative cancer specific probability of death, $\tilde{G}_{xc}$. The lower curve represents the cumulative probability that a man diagnosed with prostate cancer after the age of 70 dies from prostate cancer and the upper curve is the total cumulative probability of dying from any cause. The difference between the two curves represents the cumulative probability that a patient dies from other causes. The crude probability of death levels off over time while the net probability continues to rise, showing the substantial impact of competing risks in this population.

### *4.2. Regional breast cancer in women by age group*

The second example demonstrates how the effects of competing risks vary over different age strata. Figure 2 compares $\tilde{G}_{xc}$ and (1-relative survival) stratified by age in women diagnosed with regional breast cancer between the years 1973 and 1994. Relative survival for the different age groups is similar across age groups, confirming a common belief that stage by stage relative survival for breast cancer does not differ to a large extent in this age range. However, this does not mean that crude cause-specific probability of death does not differ by age. As shown in Figure 2, crude cumulative probability of death due to breast cancer drops off as age increases, due to hazards from other causes increasing. In this example the adjustment for $g_{xc}$ less than zero was applied to women ages 80–89 for years 13–15 after diagnosis.

Table I. Localized prostate cancer in men over the age of 70 years, 1973–1994*.

| Year | $n_x$ | $d_x$ | $l_x$ | $n_x^*$ | $P_x$ | $\Pi P_i$ | $E_x$ | $1-\Pi P_i$ | $g_{xc}$ | $g_{xc}$ (2SE) | $G_{xc}$ | $G_{xc}$ (2SE) | $g_{xo}$ | $g_{xo}$ (2SE) | $G_{xo}$ | $G_{xo}$ (2SE) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 69488 | 5916 | 4134 | 67421 | 0.9123 | 0.9123 | 0.9204 | 0.08775 | 0.008499 | 0.002273 | 0.008499 | 0.002273 | 0.07925 | $9.424 \times 10^{-5}$ | 0.07925 | $9.424 \times 10^{-5}$ |
| 2 | 59438 | 5190 | 5088 | 56894 | 0.9088 | 0.829 | 0.9165 | 0.171 | 0.007108 | 0.002303 | 0.01561 | 0.003224 | 0.07611 | 0.0002078 | 0.1554 | 0.0002938 |
| 3 | 49160 | 4955 | 5641 | 46339.5 | 0.8931 | 0.7404 | 0.9118 | 0.2596 | 0.01631 | 0.002496 | 0.03192 | 0.004031 | 0.07234 | 0.0002829 | 0.2277 | 0.0005481 |
| 4 | 38564 | 4330 | 4533 | 36297.5 | 0.8807 | 0.6521 | 0.9066 | 0.3479 | 0.02061 | 0.00265 | 0.05252 | 0.004741 | 0.06772 | 0.00035 | 0.2954 | 0.0008507 |
| 5 | 29701 | 3579 | 3074 | 28164 | 0.8729 | 0.5692 | 0.9013 | 0.4308 | 0.02011 | 0.002731 | 0.07263 | 0.005363 | 0.06275 | 0.0004116 | 0.3582 | 0.001196 |
| 6 | 23048 | 2918 | 2233 | 21931.5 | 0.8669 | 0.4935 | 0.8961 | 0.5065 | 0.01801 | 0.002764 | 0.09065 | 0.005911 | 0.05772 | 0.000467 | 0.4159 | 0.001579 |
| 7 | 17897 | 2389 | 1905 | 16944.5 | 0.859 | 0.4239 | 0.8903 | 0.5761 | 0.01689 | 0.002804 | 0.1075 | 0.006401 | 0.05269 | 0.0005163 | 0.4686 | 0.001992 |
| 8 | 13603 | 1943 | 1470 | 12868 | 0.849 | 0.3599 | 0.8844 | 0.6401 | 0.0164 | 0.002855 | 0.1239 | 0.006842 | 0.04761 | 0.0005607 | 0.5162 | 0.002431 |
| 9 | 10190 | 1515 | 1022 | 9679 | 0.8435 | 0.3036 | 0.8782 | 0.6964 | 0.01369 | 0.002847 | 0.1376 | 0.007244 | 0.04264 | 0.0006017 | 0.5588 | 0.002889 |
| 10 | 7653 | 1164 | 796 | 7255 | 0.8396 | 0.2549 | 0.8722 | 0.7451 | 0.01072 | 0.002812 | 0.1483 | 0.007615 | 0.03798 | 0.0006387 | 0.5968 | 0.003362 |
| 11 | 5693 | 932 | 665 | 5360.5 | 0.8261 | 0.2105 | 0.8654 | 0.7895 | 0.01082 | 0.002851 | 0.1592 | 0.007946 | 0.03349 | 0.00067 | 0.6303 | 0.003846 |
| 12 | 4096 | 707 | 512 | 3840 | 0.8159 | 0.1718 | 0.8582 | 0.8282 | 0.009664 | 0.00286 | 0.1688 | 0.008248 | 0.0291 | 0.0006982 | 0.6594 | 0.004333 |
| 13 | 2877 | 489 | 372 | 2691 | 0.8183 | 0.1406 | 0.8505 | 0.8594 | 0.006069 | 0.002783 | 0.1749 | 0.008559 | 0.02515 | 0.0007263 | 0.6845 | 0.004822 |
| 14 | 2016 | 361 | 286 | 1873 | 0.8073 | 0.1135 | 0.8421 | 0.8865 | 0.005492 | 0.002807 | 0.1804 | 0.008851 | 0.0216 | 0.0007504 | 0.7061 | 0.005311 |
| 15 | 1369 | 245 | 220 | 1259 | 0.8054 | 0.09139 | 0.835 | 0.9086 | 0.003688 | 0.002786 | 0.1841 | 0.009155 | 0.01839 | 0.000776 | 0.7245 | 0.005797 |

* Source: NCI's Survival Epidemiology and End Results Program. See text for description of notation.

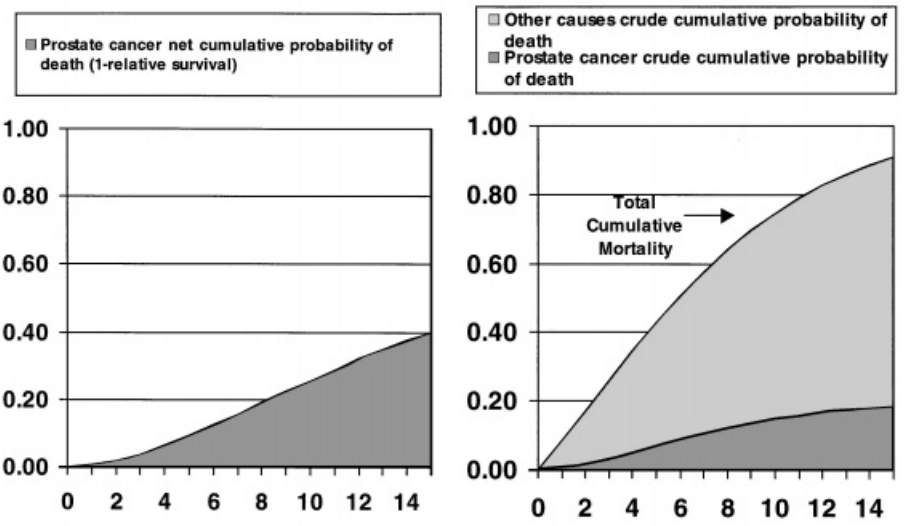


Figure 1. Cumulative probability of death in men with localized prostate cancer over the age of 70.

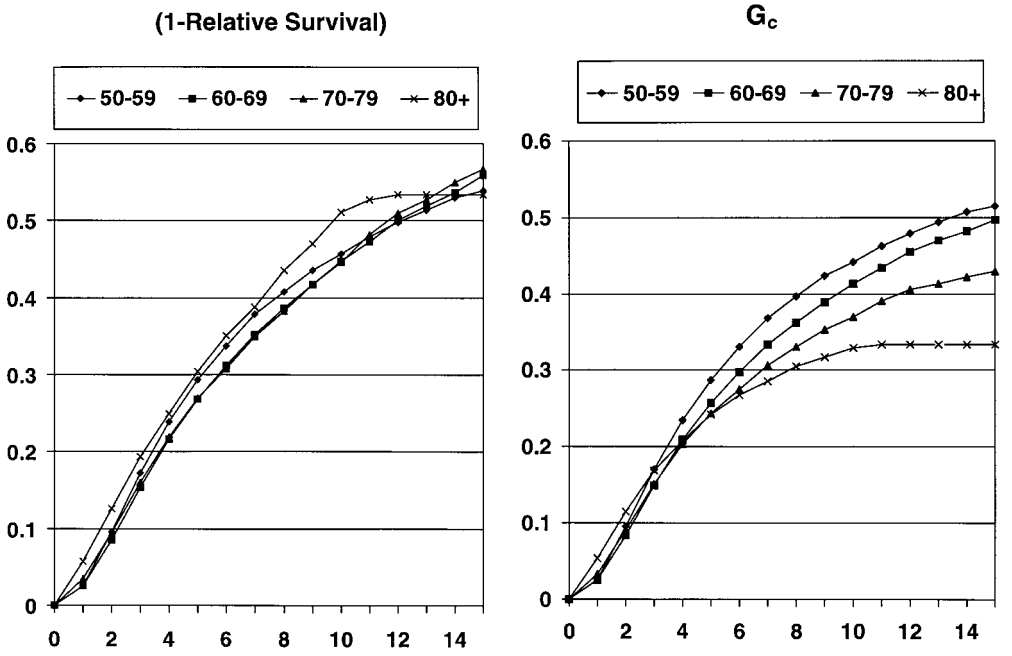


Figure 2. Cause-specific probability of death in women with regional breast cancer.

## 5. DISCUSSION

Cause-specific probability of death in the presence and absence of other causes are two distinct measures that address different questions. Net survival, estimated by relative survival, is not influenced by mortality changes in other diseases. This can be a valuable characteristic when reporting cancer progress measures. However, when assessing the impact of a cancer diagnosis at an individual level the influence of mortality from other diseases may play a key role. Available screening tests result in the diagnosis of cancers occurring earlier in the natural history of the disease, thus increasing the potential effects of competing causes of death. The example of men diagnosed with prostate cancer over the age of 70 exemplifies the situation in which crude and net measures of mortality are quite different and crude measures may be more appropriate when considering issues such as treatment options.

Estimates of relative survival and cause-specific mortality for cancer patients require the assumption of independent competing risks. Future research opportunities include investigating the effect on both net and crude mortality estimates when the assumption of independent competing risks is violated. It may be necessary to model the relationship between competing hazards for applications such as lung cancer where the risk factor of smoking influences multiple causes of death.

The need to incorporate the effects of mortality from other disease in a survival measure is reflected in other proposed cancer measures that do not rely on cause of death information. Blesch *et al.* [17] developed a measure called 'realized probability of dying (RPD)'. This measure expresses individual survival time after being diagnosed with cancer relative to the survival distribution of the reference population. The RPD is the percentage of the reference population that are still expected to be alive when the cancer patient died. For example, if the RPD is high, the cancer patient is doing much worse than the reference population since a large percentage of this population would be expected to be alive when the case dies. By looking at how much better or worse cancer patients are doing compared to the reference population, the RPD incorporates the influence of mortality from other disease.

Brown *et al.* [18] estimate the impact of cancer on population survival estimates without cause of death information. Their focus is on population mortality rates rather than case mortality as discussed in this paper. Brown *et al.* defines three measures; age adjusted proportion of the population diagnosed with cancer and projected to be dead of any cause by a particular age, age adjusted proportion of the population diagnosed and projected to be dead in excess of the overall population's experience, and expected life years lost to a 20-year-old due to the possible diagnosis of any cancer. Brown *et al.* estimated excess mortality as the difference between the survival of cancer patients and the survival of the general U.S. population conditioned on being alive at the age of diagnosis subtracted from 1.

Gooley *et al.* [19] also discuss the need to incorporate the influence of competing risks in survival measures. They consider discrete hazard models when cause of death information is known and estimate a net probability of failure (1-Kaplan–Meier) and a crude probability of failure (cumulative incidence) in the presence of competing causes of death.

Estimating cause-specific cumulative probability of death in the presence of other causes has several intuitive advantages. Most notably, it is an estimate of the mortality patterns actually observed, which would be of primary interest to cancer patients. Another intuitive advantage is the partitioning of total mortality into cause-specific categories, whereas no such partitioning is possible in the absence of other causes. Often there is interest in identifying the time point on a

survival curve where most patients could be consider cured. In the past cure models have been fit to estimates of the net survival function [20], although some authors have considered alternative definitions of cure. Haybittle [21] describes several definitions of cure, including clinical cure (patient's risk from all causes of death equal to that of the reference population) and personal cure (the latent time of death due to cancer is after the latent time of death due to other causes). Crude cause-specific probability of death is a more appropriate measure for investigating personal cure by identifying a point where the chance of death from cancer becomes negligible. In Figure 1 the probability of death from prostate cancer in the absence of other causes does not level off after 15 years. Conversely, in the presence of other causes, substantial leveling of the curve does occur by the end of the 15 year period, thus illustrating the situation where a patient's risk relative to the reference population remains elevated while the absolute risk of dying from cancer is negligible. We believe the crude probability of death due to cancer estimated using death information (for example, from cancer trials) or without cause of death information (from population-based data) is a useful supplement to the net survival function.

## APPENDIX

Assume that there are a total of $S$ strata levels ($s = 1 \ldots S$). Let $p_i^s$ equal the proportion of individuals in strata $s$ at the beginning of interval $i$, so that $p_1^s$ is the initial proportion in each strata. To properly combine the strata-specific estimates of ${}_x g_c$ we weight by the proportion in each strata at time 1, that is, $\tilde{G}_{xc} = \sum_{i=1}^{S} p_1^i \tilde{G}_{xc}^i$. In this case the estimate for $\tilde{G}_{xc}$ is

$$\sum_{s=1}^{S} p_1^s \left( \prod_{i=1}^{x-1} (1 - h_{ic}^s)(1 - {}_i h_o^s) \left( 1 - \frac{(1 - h_{xc}^s)(1 - h_{xo}^s)}{(1 - h_{xo}^s)} \right) \left( 1 - \frac{1}{2}(1 - h_{xo}^s) \right) \right) \tag{A1}$$

where $h_{xc}^s$ and $h_{xo}^s$ are the net probability of dying of cancer and other causes in strata $s$, respectively. By ignoring the different strata levels the estimate obtained is

$$\prod_{i=1}^{x-1} \left( \sum_{s=1}^{S} p_i^s (1 - h_{ic}^s)(1 - h_{io}^s) \right) \left( 1 - \frac{\sum_{s=1}^{S} p_x^s (1 - h_{xc}^s)(1 - h_{xo}^s)}{\sum_{s=1}^{S} p_x^s (1 - h_{xo}^s)} \right) \times \left( 1 - \frac{1}{2} \sum_{s=1}^{S} p_x^s (1 - h_{xo}^s) \right) \tag{A2}$$

where

$$p_i^s = \frac{p_1^s \prod_{k=1}^{i-1} ((1 - h_{kc}^s)(1 - h_{ko}^s))}{\sum_{s=1}^{S} p_1^s \prod_{k=1}^{i-1} ((1 - h_{kc}^s)(1 - h_{ko}^s))}$$

Substituting ${}_i p^s$ into the first term of equation (A2) we get

$$\left( \sum_{s=1}^{S} p_1^s \prod_{k=1}^{x-1} ((1 - h_{kc}^s)(1 - h_{ko}^s)) \right) \left( 1 - \frac{\sum_{s=1}^{S} p_x^s (1 - h_{xc}^s)(1 - h_{xo}^s)}{\sum_{s=1}^{S} p_x^s (1 - h_{xo}^s)} \right) \times \left( 1 - \frac{1}{2} \sum_{s=1}^{S} p_x^s (1 - h_{xo}^s) \right) \tag{A3}$$

It can be shown that if the net probability of dying from other causes within each interval is the same across strata (that is, $h_{io}^s = h_{io} \, \forall i$) then we have (A1) = (A3). Similarly if the net probability

of dying from cancer within each interval is the same across strata (that is, $h_{ic}^{s} = h_{ic}\ \forall i$) then again (A1) = (A3).

ACKNOWLEDGEMENTS

The authors wish to thank Dr Charles C. Brown for his helpful discussions and insightful comments on the subject during the preparation of this manuscript, and Don Green from IMS for the programming required to analyse the data shown in the two examples.

REFERENCES

1. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
2. Elandt-Johnson RC, Johnson NL. *Survival Models and Data Analysis.* Wiley: New York, 1980.
3. Cutler SJ, Ederer F, Griswold MH, Greenberg RA. Survival of breast-cancer patients in Connecticut, 1935–54. *Journal of the National Cancer Institute* 1959; **23**:1137–1156.
4. Extramural Committee to Assess Measures of Progress Against Cancer. Special report measurement of progress against cancer. *Journal of the National Cancer Institute* 1990; **82**:825–835.
5. Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *Journal of the American Statistical Association* 1993; **88**:400–409.
6. Percy CL, Miller BA, Ries LA. Effect of changes in cancer classification and accuracy of cancer death certificates on trend of cancer mortality. In *Trends in Cancer Mortality in Industrial Countries*, Davis DL, Hoel D (eds). New York Academy of Science: New York, 1990; 87–99.
7. Brown BW, Brauner C, Minnotte MC. Noncancer deaths in white adult cancer patients. *Journal of the National Cancer Institute* 1993; **85**:979–987.
8. Ederer F, Axtell LM, Cutler SJ. *The Relative Survival Rate: a Statistical Methodology*. Monograph, National Cancer Institute: Bethesda, Maryland, 1961; 101–121.
9. Ries LAG, Kosary CL, Hankey BF, Miller BA, Harras A, Edwards BK (eds). *SEER Cancer Statistics Review, 1973–1994*. National Cancer Institute, NIH Pub. No.97-2789: Bethesda, Maryland, 1997.
10. Hakulinen T, Abeywickrama KH. A computer program package for relative survival analysis. *Computer Programs in Biomedicine* 1985; **19**:197–207.
11. Cutler SJ, Ederer F. Maximum utilization of the lifetable method in analyzing survival. *Journal of Chronic Diseases* 1955; **8**:699–712.
12. Hakulinen T. Net probabilities in the theory of competing causes. *Scandinavian Actuarial Journal* 1977; 65–88.
13. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; **38**:933–942.
14. Hakulinen T. On long-term relative survival rates. *Journal of Chronic Diseases* 1977; **30**:431–443.
15. Esteve J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 1990; **9**:529–538.
16. Ries LAG, Kosary CL, Lyles LPN. Cancer patient survival: why use the relative survival rates? *The Abstract* 1995; **22**:28–30.
17. Blesch KS, Freels S, Furner S, Davis F, Miles TP. Applying the realized probability of dying to cancer survival. *Journal of Clinical Epidemiology* 1996; **49**:879–884.
18. Brown BW, Brauner C, Levy LB. Assessing changes in the impact of cancer on population survival without considering cause of death. *Journal of the National Cancer Institute* 1997; **89**:58–65.
19. Gooley TA, Leisnring W, Crowley J, Storer BE. Estimation of failure probability in the presence of competing risks: new representations of old estimators. *Statistics in Medicine* 1999; **18**:695–706.
20. Gamel JW, Meyer JS, Feuer EJ, Miller BA. The impact of stage and history on the long-term clinical course of 163,808 patients with breast carcinoma. *Cancer* 1996; **77**:1459–1464.
21. Haybittle JL. Curability of breast cancer. *British Medical Bulletin* 1990; **47**:319–323.