

## THEORY AND METHODS

# A cancer survival model that takes sociodemographic variations in "normal" mortality into account: comparison with other models

Ø Kravdal

*J Epidemiol Community Health* 2002;**56**:309–318

**Study objectives:** Sociodemographic differentials in cancer survival have occasionally been studied by using a relative survival approach, where all cause mortality among persons with a cancer diagnosis is compared with that among similar persons without such a diagnosis ("normal" mortality). One should ideally take into account that this "normal" mortality not only depends on age, sex, and period, but also various other sociodemographic variables. However, this has very rarely been done. A method that permits such variations to be considered is presented here, as an alternative to an existing technique, and is compared with a relative survival model where these variations are disregarded and two other methods that have often been used.

**Design, setting, and participants:** The focus is on how education and marital status affect the survival from 12 common cancer types among men and women aged 40–80. Four different types of hazard models are estimated, and differences between effects are compared. The data are from registers and censuses and cover the entire Norwegian population for the years 1960–1991. There are more than 100 000 deaths to cancer patients in this material.

**Main results and conclusions:** A model for registered cancer mortality among cancer patients gives results that for most, but not all, sites are very similar to those from a relative survival approach where educational or marital variations in "normal" mortality are taken into account. A relative survival approach without consideration of these sociodemographic variations in "normal" mortality gives more different results, the most extreme example being the doubling of the marital differentials in survival from prostate cancer. When neither sufficient data on cause of death nor on variations in "normal" mortality are available, one may well choose the simplest method, which is to model all cause mortality among cancer patients. There is little reason to bother with the estimation of a relative-survival model that does not allow sociodemographic variations in "normal" mortality beyond those related to age, sex, and period. Fortunately, both these less data demanding models perform well for the most aggressive cancers.

Correspondence to:  
Dr Ø Kravdal,  
Department of  
Economics, University  
of Oslo, PO Box 1095  
Blindern, N-0317  
Oslo, Norway;  
okravdal@econ.uio.no

Accepted for publication  
7 September 2001

Assessment of prognosis is a key issue in cancer epidemiology. It is important not only to check the impact that different kinds of treatment have, but also to map the overall trends in survival, and to find out whether some groups of the population fare worse than others when faced with a malignant disease. For example, a reduction of the social differentials in mortality is widely accepted as a major challenge, and appropriate interventions will require an accurate description of its sources—that is, the differentials in the incidence of and survival from various important diseases, such as the neoplasms. It would also be important, but much more difficult, to identify the mechanisms behind these differentials. Generally, cancer survival is considered to be determined by three types of factors: treatment, so called host factors, and stage at the time of diagnosis (see for example, discussion in Kravdal<sup>1,2</sup>). While the latter is taken into account in some studies (including the present), the relative importance of the two other causal channels is very hard to establish. This is regrettable, because population differences in treatment, with consequences for survival, would trigger another health policy response than survival gradients stemming from, for example, the patients' general health at the time of diagnosis.

A majority of the quite few studies of sociodemographic determinants of cancer survival suggest that being married or having a higher education or income is beneficial.<sup>1,2</sup> However, some studies point in the opposite direction, and there is no agreement on whether there are sharper survival differentials

for some cancer sites than for others. A good assessment of such differentials across sites could be an important lead in the search for explanations. For example, if the social gradients to a large extent were attributable to treatment differentials, one would expect them to be more pronounced for cancers for which an efficient treatment is reckoned to exist, than for those thought to be relatively unresponsive to treatment.

The mixed results in previous studies are partly attributable to the different data that are used, and their often small size. Besides, different countries, with more or less well developed public health care systems, have been analysed, and different methods have been used. Better knowledge about the qualities of the various methods will be of importance when assigning weights to previous studies in a review, and for choosing a good design in future research. In addition, it would, of course, be valuable to try to develop better or more convenient alternatives to existing techniques.

Two main approaches have been used in studies of how marital status, education and other sociodemographic factors affect cancer survival. The first is simply to estimate mortality differentials among people who have been diagnosed with the types of cancer under investigation, with controls for age and sex and perhaps some other potentially confounding factors (see for example, Krongrad *et al*<sup>3</sup>). This is referred to in the survival literature as *observed survival*. During the follow up period, many of these patients will die from causes that are obviously unrelated to their malignancies. In an attempt to

exclude these deaths, it is a common strategy to count only the deaths recorded as attributable to cancer and censor the observations at the time of other deaths (see for example, Neale<sup>4</sup>). This is the *corrected* or *cause specific survival*.

The other approach is based on a comparison of all cause mortality among cancer patients with that among other persons without such a diagnosis, denoted below as “normal” mortality. One advantage of this so called relative-survival approach is that data on cause of death are unnecessary. Besides, such a measure of the aggressiveness of the disease will include the excess deaths that are caused by the malignancy, but that would not (and perhaps should not) be recorded as cancer deaths. Published national mortality rates, grouped by age, sex, and period, have usually been taken as the “normal” mortality in relative survival analyses. In many studies, such measures of excess mortality have been estimated for separate groups of patients (see for example, Gilliland *et al*<sup>5</sup>), while other studies have been based on multiple regression (see for example, Schrijvers *et al*<sup>6</sup>). Special regression techniques have been developed for relative survival analysis, for example by Hakulinen and Tenkanen<sup>7</sup> and Esteve *et al*.<sup>8</sup> The former will be referred to below as the (simple) *HT-version of the relative survival* approach.

Ideally, when the goal is to assess the impact of, for example, social factors on cancer survival, also the corresponding variations in “normal” mortality should be taken into account. Such an extension was made in a study by Dickman *et al* from 1998,<sup>9</sup> where social status specific mortality rates for Finland were included as “normal” mortality within the Hakulinen-Tenkanen approach. This will be called the *extended HT-version of the relative survival* approach. For some cancer sites, the results were found to be markedly different from those obtained with a “normal” mortality depending only on age, sex, and period. In fact, the latter method was also found to be inferior to the cause specific survival approach.

Since 1996, another model that permits a broad range of sociodemographic variation in “normal” mortality has been used in cancer survival studies.<sup>1 2 10 11</sup> This model is estimated from a combined sample (CS) of cancer patients and people without such a diagnosis, and is therefore referred to below as a CS-version. To conform to the other names just given, the model can be called an *extended CS-version of the relative survival* approach when sociodemographic factors other than age, sex, and period are included in the term for “normal” mortality, and otherwise substitute the word “extended” with “simple”.

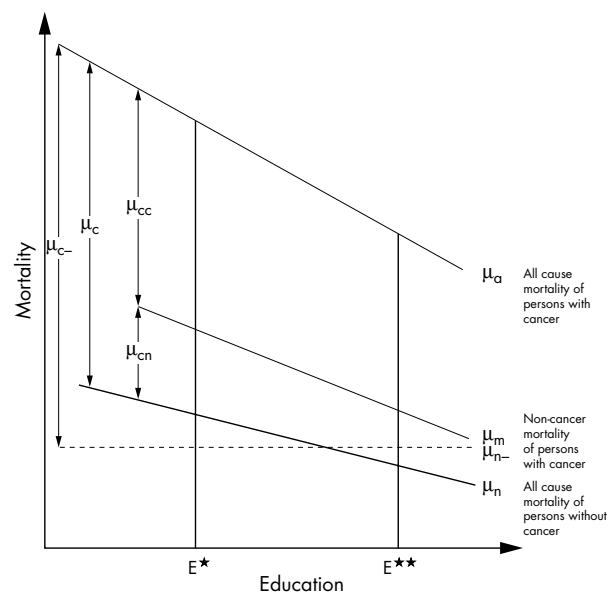
One objective of this paper is to present this alternative approach. It is not likely to produce other results than the corresponding HT-version, but it may be more convenient in some situations. The other objective is to compare it with the three other models mentioned above, that is, observed survival, cause specific survival, and relative survival without control for “normal” mortality variations except those related to age, sex, and period (the latter in its CS-version). This adds to the Finnish work referred to above. Above all, the observed survival method, which is particularly attractive because of its modest data requirements, was not included for comparison in the Finnish study. Besides, it is valuable to replicate that comparison, using other data, another technical set up (CS rather than HT), other sociodemographic variables, and partly other cancer sites, to see if the conclusions are confirmed. The focus is on education and marital status, rather than social status, as in the Finnish study.

A discussion of possible behavioural mechanisms behind the estimated effects of these variables (as well as occupation and income) can be found elsewhere.<sup>1 2</sup>

## METHODS

### Introductory issues

To set the stage for the presentation of the novel and most complex model, the basic ideas behind the relative survival



**Figure 1** A sketch of mortality by educational level for a given age, sex, and period.

approach are first reviewed. The three alternative models, which are simpler and to some extent special cases of each other, should then need little elaboration. The formal specification of all models is followed by a brief description of data and an account of some technical details. For pedagogical purposes, the differences between the models are illustrated in a graph.

A piecewise constant mortality rate is assumed in all models. With such specification, one can generally get a better fit to the duration pattern than with the less flexible parametric models. Besides, models with piecewise constant rates and categorical covariates are convenient to estimate, and are, in fact, indistinguishable from Poisson regression (of number of deaths with exposure time as “offset”, that is, conditional on exposure time).<sup>12</sup> This means that one can make use of the Poisson regression module in Epicure, which has the special advantage that it permits the partly additive structure needed for the relative survival approach.

### The rationale for the relative-survival approach

The idea behind the relative-survival approach is illustrated in figure 1, which is drawn for a given age, sex, and period. The bold line is mortality for persons without a cancer diagnosis ( $\mu_n$ ). It is depicted, for simplicity, as declining linearly with educational level, but that is of no importance for the argument that follows. According to Norwegian regulations, cancer can never be the formal cause of death among this group.

Among those who have received a cancer diagnosis, death is, of course, very often registered as the result of cancer. This “registered cancer mortality” is denoted here as  $\mu_{cr}$ , while the cancer patients’ all cause mortality is  $\mu_a$ . In addition, it is likely that persons with a cancer diagnosis may suffer an extra high non-cancer mortality as a result of their disease—that is, that their mortality from other registered causes ( $\mu_m$ ) is higher than that among other persons ( $\mu_n$ ). To give just one example, the emotional stress from cancer may increase the chance of being involved in a motor vehicle accident, which may then (correctly) be registered as the cause of death.

Because of this blurredness of the cause of death, and because cause specific mortality data are not always available, “relative survival” was suggested several decades ago as a fruitful measure of the aggressiveness of the disease. It is defined as the ratio of (all cause) survival for cancer patients

within, say, a five year interval to that among persons without such a diagnosis, and corresponds to the sum  $\mu_c$  of  $\mu_{cc}$  and  $\mu_{cn}$ . (When mortality of a cancer patient is  $\mu = \mu_n + \mu_c$ , and that of another person is  $\mu = \mu_n$ , relative survival can be found by exponentiating the negative integral of  $\mu_c$  over the relevant duration interval. Mathematically, five year relative survival is  $R_5 = \exp(-\int \mu_{cn} dw) / \exp(-\int \mu_{cn} dw) = \exp(-\int \mu_c dw)$ , where the integration is over duration  $w=0$  to  $w=5$  years.)

**The novel approach: an extended CS version of the relative survival approach (model 1a)**

The extended CS version is specified as

$$\mu = e^{b \cdot x} + y e^{c \cdot x} e^{d \cdot z}$$

where  $\mu$  is all cause mortality and  $x$  is a vector of categorical sociodemographic variables (covariates), such as age, period, educational level, and marital status.  $b$  and  $c$  are effect vectors, and  $y$  is a cancer diagnosis indicator that takes the value 1 from the time of diagnosis (if any) of the cancer under investigation, and otherwise is 0. The covariate vector  $z$  includes various characteristics of the disease, and  $d$  is the corresponding effect vector.

For a person without the cancer in focus (that is,  $y=0$ ), and with sociodemographic characteristics  $x$ , mortality is  $e^{b \cdot x}$ . This is, of course, the multiplicative structure we usually see in hazard models, and it means that mortality is a product of various positive factors. This first additive term,  $e^{b \cdot x}$ , is referred to in this study as “normal” mortality. It is close to, but not exactly equal to, the national mortality level, as people with this particular cancer, which is a quite small group, are left out. The second term,  $e^{c \cdot x} e^{d \cdot z}$ , is always positive and is the excess mortality of otherwise similar persons who have been diagnosed with this particular type of cancer.

**The three simpler models used in previous studies**

**Model 1b: the simple CS version of the relative survival approach**

The inclusion of only age, sex, and period in the first term, and not education or marital status, can be symbolised by a different covariate  $x$ - in the equation, which thus looks like:

$$\mu = e^{b' \cdot x} + y e^{c' \cdot x} e^{d' \cdot z}$$

This means that “normal” mortality (given age, sex, and period) is assumed to be a constant line  $\mu_n$ , rather than a varying  $\mu_n$ .

This model is essentially the same as the simple HT version of the relative survival approach, except that “normal” mortality is modelled, rather than taken as constants.

**Model 2a: observed survival**

The simplest approach is to model all cause mortality for a sample of cancer patients exclusively. This model can be written as

$$\mu_{cc} = e^{g \cdot x} e^{h \cdot z}$$

**Model 2b: cause specific survival**

Alternatively, one may estimate cancer patients’ risk of dying from cancer, according to the “main cause” of death recorded on the death certificates. The model thus becomes

$$\mu_{cc} = e^{g \cdot x} e^{h \cdot z}$$

In the Norwegian data, a large majority of those who die within five years of diagnosis are reported as dying from the malignant disease. The proportion is 70%–95%, depending on the cancer site, with 90% as an average.

**Table 1** Estimates (with 95% confidence intervals) of differentials in “normal” mortality and excess mortality from colorectal cancer among Norwegian men, according to register and census data for 1960–1991

	Number of deaths	Estimates
<i>“Normal” mortality</i>		
Constant		0.0025 deaths per person per month
Educational level		
7–9 years†	282421	1
10–12	80673	0.86* (0.85 to 0.86)
13–16	17114	0.75* (0.74 to 0.77)
17–	6521	0.64* (0.62 to 0.66)
Age		
50–54	22402	0.24* (0.23 to 0.24)
55–59	35849	0.39* (0.38 to 0.39)
60–64	54194	0.63* (0.62 to 0.63)
65–69†	75718	1
70–74	94996	1.60* (1.59 to 1.62)
75–79	103570	2.53* (2.51 to 2.55)
Period		
1960–69†	111569	1
1970–79	130520	1.03* (1.02 to 1.03)
1980–85	66898	0.99* (0.98 to 1.00)
1986–91	77742	0.95* (0.94 to 0.96)
<i>Excess mortality</i>		
Constant		0.0051 deaths per person per month‡
Educational level		
7–9 years†	7089	1
10–12	2555	0.95* (0.90 to 1.00)
13–16	665	0.87* (0.80 to 0.96)
17–	292	0.83* (0.73 to 0.95)
Age		
50–54	539	0.95 (0.86 to 1.05)
55–59	921	0.93 (0.86 to 1.01)
60–64	1496	0.96 (0.89 to 1.03)
65–69†	2194	1
70–74	2660	1.09* (1.02 to 1.16)
75–79	2791	1.19* (1.11 to 1.27)
Period		
1960–69†	2284	1
1970–79	3376	0.80* (0.75 to 0.85)
1980–85	2161	0.65* (0.61 to 0.70)
1986–91	2780	0.65* (0.60 to 0.69)
Subsite		
caecum, ascendance†	1789	1
transverse, descendance	1412	1.09* (1.01 to 1.18)
sigmoideum	2839	1.03 (0.96 to 1.10)
rectum	4561	1.27* (1.19 to 1.35)
Stage		
local†	2630	1
regional	3735	2.42* (2.26 to 2.59)
distant	4080	12.61* (11.82 to 13.44)
unknown	156	5.24* (4.39 to 6.28)
Histological type/grade		
adenocarcinoma, well differentiated†	379	1
adenocarc., moderately differentiated	1911	1.20* (1.05 to 1.39)
(adeno) carc., poorly diff.	830	2.37* (2.04 to 2.75)
other	7481	1.50* (1.31 to 1.71)

†Arbitrarily chosen reference category; ‡This excess mortality, which is experienced for men in the reference categories, corresponds to a five year relative survival of 74%; \*significant at the 0.05 level.

**Data**

The analysis is based on individual sociodemographic biographies for all men and women with a Norwegian identification number (that is, all those who have lived in Norway for some time after 1960). These life histories have been extracted from the Norwegian Population Register, the Cause of Mortality Register and the Population Censuses of 1960, 1970, and 1980, and include information about date and cause of death and all changes in residence, as well as marital status, education, income and occupation at the time of the censuses. They have

**Table 2** Estimates (with 95% confidence intervals) of educational differentials in excess mortality from cancer among Norwegian men and women, by cancer site, according to register and census data for 1960–1991†

Educational level (years)	Women		Men	
	Number of deaths	Estimates	Number of deaths	Estimates
<b>Stomach‡</b>				
7–9	5361	1	8809	1
10–12	1079	1.00 (0.93 to 1.07)	2113	0.95* (0.90 to 0.99)
13–16	135	0.96 (0.80 to 1.14)	408	0.94 (0.84 to 1.04)
17–			128	0.85 (0.71 to 1.03)
<b>Colon / rectum§</b>				
7–9	6923	1	7089	1
10–12	1871	0.87* (0.82 to 0.92)	2555	0.95* (0.90 to 1.00)
13–16	352	0.89 (0.80 to 1.01)	665	0.87* (0.80 to 0.96)
17–			292	0.83* (0.73 to 0.95)
<b>Pancreas¶</b>				
7–9	2597	1	3105	1
10–12	707	1.04 (0.95 to 1.13)	1029	0.93* (0.87 to 1.00)
13–16	119	0.79* (0.65 to 0.95)	256	0.80* (0.71 to 0.92)
17–			85	1.04 (0.83 to 1.30)
<b>Lung**</b>				
7–9	2798	1	11252	1
10–12	961	1.04 (0.96 to 1.12)	3635	0.93* (0.89 to 0.97)
13–16	185	1.15 (0.99 to 1.34)	705	0.87* (0.81 to 0.94)
17–			254	0.82* (0.72 to 0.93)
<b>Breast††</b>				
7–9	8170	1		
10–12	2721	0.91* (0.86 to 0.96)		
13–	548	0.76* (0.69 to 0.84)		
<b>Cervix‡‡</b>				
7–9	2650	1		
10–12	624	1.05 (0.95 to 1.16)		
13–	89	0.91 (0.72 to 1.14)		
<b>Corpus uteri§§</b>				
7–9	1766	1		
10–12	507	0.96 (0.95 to 1.08)		
13–	95	0.89 (0.69 to 1.13)		
<b>Ovaries¶¶</b>				
7–9	4540	1		
10–12	1538	1.00 (0.93 to 1.06)		
13–	298	0.91 (0.81 to 1.03)		
<b>Prostate***</b>				
7–9			9310	1
10–12			3054	0.90* (0.85 to 0.95)
13–16			693	0.96 (0.87 to 1.06)
17–			314	0.66* (0.56 to 0.76)
<b>Bladder†††</b>				
7–9	1176	1	3018	1
10–12	252	0.83* (0.71 to 0.97)	1094	0.89* (0.81 to 0.97)
13–16	47	0.70* (0.50 to 0.98)	215	0.79* (0.66 to 0.94)
17–			91	0.59* (0.44 to 0.80)
<b>Malignant melanoma‡‡‡</b>				
7–9	606	1	895	1
10–12	298	1.01 (0.85 to 1.20)	487	0.72* (0.63 to 0.83)
13–16	71	0.77 (0.56 to 1.05)	153	0.58* (0.47 to 0.72)
17–			63	0.60* (0.44 to 0.83)
<b>Leukaemia§§§</b>				
7–9	2518	1	3091	1
10–12	773	0.94 (0.86 to 1.02)	1093	0.90* (0.83 to 0.97)
13–16	121	0.84 (0.69 to 1.02)	261	0.81* (0.71 to 0.93)
17–			117	0.72* (0.59 to 0.88)
<b>All sites above¶¶¶</b>				
7–9	38395	1	46323	1
10–12	11198	0.96* (0.94 to 0.98)	14953	0.92* (0.90 to 0.93)
13–16	2037	0.86* (0.82 to 0.91)	3757	0.88* (0.84 to 0.91)
17–			1316	0.78* (0.73 to 0.83)

Footnotes on next page

been linked with data from the Norwegian Cancer Registry, which from 1953 has received information on all cancer cases in the population (site, basis for the diagnosis, histological grade and type, and the stage of the disease at the time of diagnosis). These linked data have been used in several previous studies.<sup>1 2 10 11 13</sup>

Twelve common cancer sites are considered in the analysis. At ages 40–80 during the observation period 1960–1991, there were more than 150 000 deaths among men or women

diagnosed with cancer less than five years previously. Three quarters of these deaths were among persons with cancer in one of the 12 selected sites.

#### Detailed specifications

In all these models, the men and women are followed up from age 50 (or 40 in models for leukaemia, malignant melanoma, and cancer in the breast or female genitals, which tend to occur at a lower age) or, if born before 1910 (or 1920), from

**Table 2** continued from previous page

†The models also included age, period, subsite, stage, and histological type/grade (the latter three only in the excess mortality term). ‡The subsite control variable had four categories: (i) pars pylorica, canal, angulus, antrum, corpus, (ii) previously recessed due to benign ulcer, (iii) several locations, and not specified, (iv) all other. The histological control variable had four categories: (i) well differentiated adenocarcinoma, (ii) moderately well differentiated adenocarcinoma, (iii) poorly differentiated adenocarcinoma or carcinoma, (iv) all other. §The subsite control variable had four categories: (i) caecum and colon ascendance, (ii) colon transversum and colon descendance, (iii) sigmoideum, (iv) rectum. The few other colon subsites and the anal channel were excluded. The histological control variable had four categories: (i) well differentiated adenocarcinoma, (ii) moderately well differentiated adenocarcinoma, (iii) poorly differentiated adenocarcinoma or carcinoma, (iv) all other. ¶The subsite control variable had three categories: (i) caput (head), (ii) pancreas not specified, (iii) all other. The histological control variable had four categories: (i) well differentiated adenocarcinoma, (ii) moderately well differentiated adenocarcinoma, (iii) poorly differentiated adenocarcinoma or carcinoma, (iv) all other. \*\*Two histological control variables were included. One of them was grade: (i) well differentiated, (ii) moderately well differentiated, (iii) poorly differentiated, (iv) all other. The other was type: (i) squamous cell carcinoma, (ii) small cell carcinoma, (iii) adenocarcinoma, (iv) all other. ††The histological control variable had four categories, referring only to grade: (i) well differentiated, (ii) moderately well differentiated, (iii) poorly differentiated, (iv) all other. †††Two histological control variables were included. One of them was grade: (i) well differentiated, (ii) moderately well differentiated, (iii) poorly differentiated, (iv) all other. The other was type (i) adenocarcinoma, (ii) squamous cell carcinoma, (iii) all other. §§Two histological control variables were included. One of them was grade: (i) well differentiated, (ii) moderately well differentiated, (iii) poorly differentiated, (iv) all other. The other was type (i) adenocarcinoma, (ii) adenosquamous carcinoma, adenocanthoma, (iii) leiomyoma, (iv) all other. ¶¶Two histological control variables were included. One of them was grade: (i) well differentiated, (ii) moderately well differentiated, (iii) poorly differentiated, (iv) all other. The other was type (i) adenocarcinoma, (ii) serous cystadenocarcinoma, (iii) mucinost cystadenoma, (iv) all other. \*\*\*The histological control variable had four categories: (i) well differentiated adenocarcinoma or carcinoma, (ii) moderately well differentiated adenocarcinoma or carcinoma, (iii) poorly differentiated adenocarcinoma or carcinoma, (iv) all other. ††††The few cancers in urethra were excluded. The histological control variable had four categories: (i) slightly atypical transitional and papillary tumor or (in older classification) well differentiated adenocarcinoma or squamous cell carcinoma, (ii) somewhat atypical transitional or papillary tumor or (in older classification) moderately well differentiated adenocarcinoma or squamous cell carcinoma, (iii) markedly atypical transitional or papillary tumor or (in older classification) poorly differentiated adenocarcinoma or squamous cell carcinoma, (iv) all other. †††††The subsite control variable had four categories: (i) head and neck, (ii) body excluding breast, (iii) upper extremities/lower extremities except feet, (iii) all other. The histological control variable had four categories: (i) malignant melanoma with superficial spread, invasive, (ii) nodular malignant melanoma, (iii) lentigo malignant melanoma, (iv) all other. §§§The histology/subsite variable had seven categories: (i) myelomatosis, (ii) chronic lymphocytic leukaemia, (iii) chronic myelocytic leukaemia, (iv) acute leukaemia/blast leukaemia, (v) acute lymphocytic leukaemia, (vi) acute myelocytic leukaemia, (vii) all other leukaemia. Stage was not a relevant variable. ¶¶¶This particular model also included age, period, stage, and site (the latter two, when relevant, only in the excess mortality term). \*Significant at the 0.05 level.

1960. It is censored at age 80 (because of potentially inadequate case ascertainment at higher ages), at the time of emigration, or by the end of 1991, which was the end of the follow up period. Moreover, people are excluded from the analysis (that is, contribute neither exposure nor deaths) during periods when they lived in another country. In addition, those who did not participate in one of the censuses are excluded during the 10 years after this census. These two restrictions have very little influence on total exposure time. The very few persons with a cancer diagnosed at necropsy are treated as having never received a cancer diagnosis. (Estimated survival gradients might be slightly affected, because of possible sociodemographic differentials in the diagnostic intensity or necropsy frequency, but this should be of no concern when the focus is on differences between models).

Some cancer types may cause death already after a few months, while others may be relatively harmless for many years, perhaps followed by recurrences. In this study, the observations are censored five years after diagnosis, if any. This interval is commonly used in cancer survival analysis. The model thus provides information on how the various sociodemographic variables influence excess mortality “on the whole” over the five years after diagnosis. (A more detailed picture could have been established by including interactions with duration since diagnosis, or by estimating additional models for other interval lengths.)

Mortality is assumed to be constant within five year age groups, which was experimentally found to be a sufficient control for age.

There is no distinction between persons who had previously been diagnosed with another type of cancer and those for whom the cancer in focus was the first one.

All covariates are categorical. They are also time varying: a level for the covariate is defined for each month during the follow up period, and refers to the situation at that time (age, period) or that in the last previous census (education and marital status).

The  $\mathbf{z}$ -vector includes subsite, stage, and histological type and/or grade, whenever relevant. Stage is defined as localised, regional spread, distant spread, or unknown. The categorisation of histology and subsite differs across cancer sites and is explained in notes to table 2. The intention behind the categorisation is to distinguish at least between quite large

histological groups or subsites associated with markedly different survival rates.

In addition, a few models that are not site specific are estimated. A site variable is then included in  $\mathbf{z}$ , and (for the models 1a and 1b)  $y$  takes the value 1 from the time of diagnosis of any of the 12 cancers. Only the first cancer diagnosis is considered in this set up.

The models are estimated in the Amfit Poisson regression module in the Epicure software.<sup>14</sup> A self made computer program (written in Pascal), operating on the individual level register and census data, was used to compute the tables of deaths and exposures that were fed into Amfit. For each person and each month of follow up, a contribution of one month was added in the appropriate cell of the exposure table (defined according to the level of each covariate at that time). One death was added to the table of deaths, in the appropriate cell, if follow up was ended by death.

## RESULTS

### Education effects estimated from the model that permits educational variations in “normal” mortality

For pedagogical purposes, all estimates from model 1a, including those for variations in “normal” mortality, are shown for one cancer site. Colorectal cancer has been chosen (table 1). Men in the arbitrarily chosen reference category (that is, aged 65–69, observed 1960–69, compulsory education) have a “normal” mortality of 0.0025 deaths per person month. Otherwise similar men who have a post-secondary education at a low level have a “normal” mortality that is 25% lower, while it is 36% lower for the few who have reached a level corresponding to a master’s degree. (The corresponding figures for women are almost the same; not shown).

Turning now to the second term, it is found that men in the reference category who have a localised, well differentiated adenocarcinoma in caecum or ascendance (which is the reference category for the disease variable) experience 0.0051 deaths per person month more than otherwise similar men without this disease. For those who have an education at a master’s level, but are otherwise similar, the corresponding excess mortality is 17% lower than this. The excess mortality declines monotonically with education.

The estimated effects of educational level on excess mortality from colorectal cancer in men are shown also in

**Table 3** Estimates of educational differentials in excess mortality from cancer among Norwegian men and women, by cancer site, according to different methods, 1960–1991†

	Models for women				Models for men			
	1a	1b	2a	2b	1a	1b	2a	2b
Stomach								
10–12	1.00	0.99	0.99	1.00	0.95*	0.94*	0.94*	0.94*
13–16	0.96	0.95	0.94	0.95	0.94	0.92	0.92*	<b>0.91</b>
17–					0.85	0.83*	0.84*	0.84
Colon/rectum								
10–12	0.87*	0.86*	0.86*	0.87*	0.95*	0.93*	0.94*	0.93*
13–16	0.89	0.87*	0.87*	0.90	0.87*	<b>0.84*</b>	0.85*	0.86*
17–					0.83*	<b>0.78*</b>	<b>0.80*</b>	0.83*
Pancreas								
10–12	1.04	1.03	1.03	1.03	0.93*	0.93*	0.93*	0.95
13–16	0.79*	0.78*	0.79*	<b>0.76*</b>	0.80*	0.80*	0.80*	0.79*
17–					1.04	1.03	1.03	1.04
Lung								
10–12	1.04	1.03	1.03	1.03	0.93*	0.92*	0.93*	0.93*
13–16	1.15	1.14	1.14	1.13	0.87*	0.86*	0.87*	0.86*
17–					0.82*	0.80*	0.81*	0.80*
Breast								
10–12	0.91*	<b>0.88*</b>	0.89*	0.90*				
13–	0.76*	<b>0.73*</b>	0.75*	0.76*				
Cervix								
10–12	1.05	1.03	<b>1.02</b>	1.05				
13–	0.91	0.89	<b>0.87</b>	<b>0.96</b>				
Corpus uteri								
10–12	0.96	<b>0.93</b>	<b>0.92</b>	0.98				
13–	0.89	<b>0.84</b>	<b>0.84</b>	<b>0.92</b>				
Ovaries								
10–12	1.00	0.99	0.99	0.99				
13–	0.91	0.90	0.90	0.89				
Prostate								
10–12					0.90*	<b>0.86*</b>	0.89*	0.91*
13–16					0.96	<b>0.89*</b>	<b>0.90*</b>	0.94*
17–					0.66*	<b>0.56*</b>	0.65*	<b>0.70*</b>
Bladder								
10–12	0.83*	0.81*	0.83*	0.81*	0.89*	<b>0.85*</b>	0.88*	0.87*
13–16	0.70*	<b>0.66*</b>	0.72*	0.69*	0.79*	<b>0.74*</b>	<b>0.76*</b>	0.79*
17–					0.59*	<b>0.49*</b>	<b>0.62*</b>	<b>0.64*</b>
Malignant melanoma								
10–12	1.01	<b>0.97</b>	<b>0.95</b>	<b>0.94</b>	0.72*	0.70*	<b>0.75*</b>	0.72*
13–16	0.77	<b>0.71*</b>	0.78*	<b>0.69</b>	0.58*	<b>0.55*</b>	<b>0.62*</b>	<b>0.62*</b>
17–					0.60*	<b>0.55*</b>	0.59*	<b>0.63*</b>
Leukaemia								
10–12	0.94	0.93	0.93	0.92	0.90*	0.89*	0.90*	0.89*
13–16	0.84	0.84	0.84	0.84	0.81*	0.80*	0.81*	0.82*
17–					0.72*	0.71*	0.71*	0.71*
All sites above								
10–12	0.96*	0.94*	0.94*	0.95*	0.92*	0.90*	0.91*	0.91*
13–16	0.86*	0.84*	0.85*	0.86*	0.88*	<b>0.85*</b>	0.86*	0.87*
17–					0.78*	<b>0.73*</b>	<b>0.75*</b>	0.78*

†The models also included age, period, subsite, stage, and histological type/grade (the latter three, when relevant, only in the excess mortality term). The subsite and histological control variables were as defined in table 2. Only effects relative to those with 7–9 years of education are shown. Differences of 0.03–0.05, when compared with results from model 1a, are in bold types, and those larger than 0.05 are in bold, underlined types. \*Significant at the 0.05 level.

table 2, along with corresponding estimates from other cancer sites and for women. The figures at the bottom of the table are for all the 12 cancer sites pooled.

#### Education effects estimated from other models

In table 3, the estimates shown in table 2 are compared with those from the three other models. For simplicity, men or women with only compulsory education, which are the reference group, are excluded from this table. (To facilitate the reading of the table, differences of 0.03–0.05 are shown in bold types and those larger than 0.05 in underlined bold types.)

When “normal” mortality is assumed to be independent of education (model 1b), the excess mortality of those with secondary or higher education compared with those with only a compulsory education becomes smaller. In other words, the effects of education on survival become sharper.

The differences are minor for some malignancies, such as lung cancer, pancreas cancer and stomach cancer, but are quite large for prostate cancer, bladder cancer and malignant melanoma. According to model 1b, men with an education at the master's level would have an excess mortality from prostate cancer that is 44% lower than that among men with only compulsory education. However, when it is taken into account that they would have had a much lower mortality also in the absence of the disease (model 1a), this advantage is reduced to 34%. The corresponding figures for men with one to four years of secondary education are 11% and 4%.

Because the control for educational differentials in “normal” mortality has little impact for most of the cancer sites, the differences between models 1a and 1b are rather small when all sites are pooled. They range from 0.02 to 0.05.

**Table 4** Estimates (with 95% confidence intervals) of marital differentials in excess

Marital status	Women		Men	
	Number of deaths	Estimates	Number of deaths	Estimates
Stomach				
never married	991	1.09* (1.02 to 1.18)	1285	1.14* (1.07 to 1.21)
married†	3734	1	8964	1
divorced/separated	226	0.98 (0.85 to 1.13)	336	1.15* (1.02 to 1.29)
widowed	1624	1.01 (0.95 to 1.08)	873	1.10* (1.01 to 1.18)
Colon/rectum				
never married	1341	1.17* (1.09 to 1.25)	1130	1.27* (1.18 to 1.36)
married†	5434	1	8391	1
divorced/separated	356	0.96 (0.85 to 1.08)	343	1.00 (0.87 to 1.14)
widowed	2015	1.01 (0.95 to 1.07)	737	1.06 (0.96 to 1.16)
Pancreas				
never married	401	1.16* (1.04, 1.30)	475	1.05 (0.95 to 1.15)
married†	2033	1	3516	1
divorced/separated	162	1.18* (1.00 to 1.39)	192	1.33* (1.14 to 1.54)
widowed	825	0.95 (0.87 to 1.04)	292	1.08 (0.95 to 1.23)
Lung				
never married	402	1.14* (1.02 to 1.27)	1473	1.24* (1.17 to 1.31)
married†	2373	1	12562	1
divorced/separated	361	1.17* (1.05 to 1.31)	920	1.18* (1.10 to 1.26)
widowed	808	1.19* (1.09 to 1.30)	927	1.13* (1.05 to 1.21)
Breast				
never married	1907	1.11* (1.05 to 1.18)		
married†	7161	1		
divorced/separated	443	1.06 (0.94 to 1.18)		
widowed	1928	1.06 (0.99 to 1.13)		
Cervix				
never married	341	1.27* (1.12 to 1.44)		
married†	2093	1		
divorced/separated	369	1.26* (1.12 to 1.43)		
widowed	560	1.04 (0.92 to 1.16)		
Corpus uteri				
never married	444	1.04 (0.92 to 1.19)		
married†	1387	1		
divorced/separated	74	0.90 (0.68 to 1.21)		
widowed	463	0.99 (0.86 to 1.13)		
Ovaries				
never married	983	1.09* (1.01 to 1.18)		
married†	4178	1		
divorced/separated	253	1.03 (0.90 to 1.18)		
widowed	962	1.01 (0.93 to 1.09)		
Prostate				
never married			1317	1.14* (1.06 to 1.24)
married†			10464	1
divorced/separated			430	1.23* (1.07 to 1.41)
widowed			1151	1.00 (0.91 to 1.10)
Bladder				
never married	237	0.89 (0.75 to 1.05)	489	1.53* (1.36 to 1.72)
married†	789	1	3449	1
divorced/separated	88	1.09 (0.85 to 1.41)	153	1.18 (0.94 to 1.48)
widowed	415	1.04 (0.90 to 1.20)	327	1.31* (1.12 to 1.53)
Malignant melanoma				
never married	127	0.97 (0.76 to 1.24)	178	1.65* (1.37 to 1.98)
married†	667	1	1312	1
divorced/separated	40	0.74 (0.47 to 1.16)	54	1.09 (0.77 to 1.53)
widowed	141	0.79 (0.60 to 1.03)	54	1.22 (0.83 to 1.80)
Leukaemia				
never married	452	1.05 (0.94 to 1.18)	454	1.11 (0.99 to 1.27)
married†	2144	1	3698	1
divorced/separated	100	0.90 (0.72 to 1.12)	114	1.07 (0.86 to 1.32)
widowed	716	1.09 (0.99 to 1.20)	296	1.07 (0.93 to 1.23)
All sites above				
never married	7501	1.14* (1.11 to 1.17)	6927	1.20* (1.17 to 1.24)
married†	31693	1	51946	1
divorced/separated	2422	1.07* (1.03 to 1.12)	2592	1.19* (1.14 to 1.24)
widowed	10014	1.06* (1.03 to 1.09)	4484	1.09* (1.05 to 1.13)

†The models also included age, period, subsite, stage, and histological type/grade (the latter three, when relevant, only in the excess mortality term). The subsite and histological control variables were as defined in table 2. \*Significant at the 0.05 level.

While the exclusion of education from the first term of the model reduces the estimated excess mortality relative to the reference category (that is, gives an exaggerated impression of the importance of educational level for survival), there is a further change in the opposite direction when the term is

entirely removed (model 2a). With only quite few exceptions, the estimates are closer to 1 than those obtained from model 1b, and thus also nearer those from model 1a.

If only the deaths reported as attributable to the cancers in focus are counted (model 2b), the effects of education are, on

**Table 5** Estimates of marital differentials in excess mortality from cancer among Norwegian men and women, by cancer site, according to different methods, 1960–1991†

	Models for women				Models for men			
	1a	1b	2a	2b	1a	1b	2a	2b
Stomach								
never married	1.09*	1.10*	1.10*	1.10*	1.14*	1.16*	1.15*	1.15*
divorced/separated	0.98	0.99	0.99	0.96	1.15*	<b>1.19*</b>	<b>1.18*</b>	1.14*
widowed	1.01	1.02	1.02	1.02	1.10*	1.11*	1.11*	1.12*
Colon/rectum								
never married	1.17*	1.19*	1.18*	1.18*	1.27*	<b>1.31*</b>	1.28*	<b>1.31*</b>
divorced/separated	0.96	<b>0.99</b>	0.98	0.95	1.00	<b>1.10</b>	<b>1.10</b>	<b>1.11</b>
widowed	1.01	1.02	1.03	1.01	1.06	<b>1.11*</b>	1.08*	<b>1.10*</b>
Pancreas								
never married	1.16*	1.16*	1.16*	1.15*	1.05	1.05	1.05	1.05
divorced/separated	1.18*	1.17*	1.17*	1.17	1.33*	1.34*	1.34*	<b>1.26*</b>
widowed	0.95	0.95	0.95	0.95	1.08	1.09	1.09	1.11
Lung								
never married	1.14*	1.14*	1.14*	1.14*	1.24*	1.25*	1.25*	1.25*
divorced/separated	1.17*	1.18*	1.18*	1.18*	1.18*	1.20*	1.20*	1.19*
widowed	1.19*	1.19*	1.19*	1.18*	1.13*	1.14*	1.13*	1.12*
Breast								
never married	1.11*	<b>1.15*</b>	1.12*	<b>1.15*</b>				
divorced/separated	1.06	<b>1.11*</b>	<b>1.11*</b>	1.05				
widowed	1.06	<b>1.09*</b>	1.07*	<b>1.10*</b>				
Cervix								
never married	1.27*	1.29*	1.26*	1.29*				
divorced/separated	1.26*	<b>1.30*</b>	1.28*	1.26*				
widowed	1.04	1.05	1.05	1.05				
Corpus uteri								
never married	1.04	1.07	1.06	1.05				
divorced/separated	0.90	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>				
widowed	0.99	<b>1.02</b>	<b>1.04</b>	1.01				
Ovaries								
never married	1.09*	1.10*	1.10*	1.10*				
divorced/separated	1.03	1.05	1.05	1.04				
widowed	1.01	1.01	1.02	1.00				
Prostate								
never married					1.14*	<b>1.24*</b>	<b>1.19*</b>	<b>1.20*</b>
divorced/separated					1.23*	<b>1.45*</b>	<b>1.34*</b>	<b>1.20</b>
widowed					1.00	<b>1.09*</b>	<b>1.07*</b>	<b>1.06</b>
Bladder								
never married	0.89	0.91	<b>0.93*</b>	0.91	1.53*	<b>1.62*</b>	<b>1.47*</b>	1.51*
divorced/separated	1.09	<b>1.13</b>	<b>1.15</b>	<b>1.03</b>	1.18	<b>1.38*</b>	<b>1.29*</b>	1.20*
widowed	1.04	1.06	1.06	1.05	1.31*	<b>1.40*</b>	<b>1.25*</b>	<b>1.35*</b>
Malignant melanoma								
never married	0.97	<b>1.01</b>	<b>1.00</b>	<b>1.04</b>	1.65*	<b>1.71*</b>	<b>1.56*</b>	<b>1.69*</b>
divorced/separated	0.74	<b>0.81</b>	<b>1.01</b>	<b>0.69</b>	1.09	<b>1.21</b>	<b>1.20</b>	<b>1.17</b>
widowed	0.79	<b>0.82</b>	<b>0.87</b>	<b>0.95</b>	1.22	<b>1.32</b>	<b>1.16</b>	<b>1.18</b>
Leukaemia								
never married	1.05	1.06	1.05	<b>1.09</b>	1.11	<b>1.14*</b>	1.13*	1.12*
divorced/separated	0.90	<b>0.94</b>	<b>0.93</b>	0.91	1.07	<b>1.16</b>	<b>1.12</b>	1.06
widowed	1.09	1.11*	1.09	1.11	1.07	1.09	1.09	<b>1.10</b>
All sites above								
never married	1.14*	1.16*	1.14*	1.16*	1.20*	<b>1.24*</b>	<b>1.23*</b>	1.22*
divorced/separate	1.07*	<b>1.10*</b>	1.09*	1.06*	1.19*	<b>1.26*</b>	<b>1.24*</b>	1.20*
widowed	1.06*	1.07*	1.06*	1.07*	1.09*	<b>1.12*</b>	1.11*	1.11*

†The models also included age, period, subsite, stage, and histological type/grade (the latter three, when relevant, only in the excess mortality term). Only effects relative to the married are shown. The subsite and histological control variables were as defined in table 2. Differences of 0.03–0.05, when compared with results from model 1a, are in bold type, and those larger than 0.05 are in bold, underlined type. \*Significant at the 0.05 level.

the whole, slightly weaker. When all cancer sites are pooled, the differences between model 2a and model 2b are 0.01–0.03.

### Marital status effects

Similar patterns appear when marital status is considered instead of education. The effects of marital status on excess mortality from different types of cancer, according to the most complex model (1a), are shown in table 4.

The differences in “normal” mortality are even larger for marital status than for education. Never married men have a 30% higher ‘normal’ mortality than the married, while that of widowers is 25% higher and that of the divorced 79% higher (not shown). The corresponding figures for women are lower.

Divorced men are found to have an excess mortality from prostate cancer that is 45% higher than that of married men, according to model 1b, but when their much higher mortality in the absence of the disease is taken into account (model 1a), this difference is reduced to 23% (table 5). Similarly, the relatively high excess mortality for widowers completely disappears, and that for never married is almost halved, from 24% to 14%.

Quite large differences between models are found also for colorectal cancer, bladder cancer, and malignant melanoma. The differences between models 1b and 2a are as explained above for models including educational level: by and large,



estimates from model 2a are nearer 1 and nearer those from model 1a.

If registered cancer mortality is modelled instead (model 2b), the effects are, on the whole, slightly weaker than those from model 2a. When all cancer sites are pooled, the differences between model 2a and model 2b for divorcees are 0.03–0.04. The differences are smaller for the never married and widowed.

## DISCUSSION

### A view to the Hakulinen-Tenkanen version

In the HT version of the relative survival approach, “normal” mortality (or, rather, survival) rates are defined for relevant subgroups, for example all possible combinations of age, sex, and period, and considered as constants. A ratio between patient survival and “normal” (often denoted as “expected”) survival is calculated for each covariate combination (cell). Using a log-log transformation of this ratio, which yields a linear structure of sociodemographic (and other) effects, the estimation is easily carried out in the GLIM software.

In the extended version of this,<sup>9</sup> social-status-specific “normal” mortality rates were first calculated from individual data (but, as explained in the paper, they could also have been taken from published life tables adjusted by social status correction factors derived from other sources). However, if such individual data exist, and can be linked with data on cancer diagnosis, also model 1a presented here can be estimated. Model 1a should in some respects be slightly more satisfying, without necessarily giving perceptibly different results.

Firstly, “normal” mortality is not treated as a constant. When data for a large national population are used to establish “normal” mortality, it is, of course, acceptable to use only point estimates for each cell and disregard information on variance. This is less satisfying when one has access only to a smaller sample (perhaps supposed to be representative of the national population). When variances in “normal” mortality are no longer negligible compared with the variances in patient mortality, disregarding the variation between cells in these variances (by essentially setting them all to 0) will, in principle, affect even point estimates of effects.

Secondly, “normal” mortality in model 1a does not include the mortality among the cancer patients in focus, which may be a non-negligible contribution for some very common and aggressive cancer types.

Thirdly, the log-log transformation in the HT version requires that cells with no surviving cancer patients, which may be found especially when the patient sample is small, must be excluded from the analysis. This could have an effect on estimates.

Fourthly, it is practically convenient to have a one step approach rather than an approach that requires first the estimation of “normal” mortality rates, which are then used in a second step to estimate gradients in relative survival. On the other hand, such an argument may count little compared with software preferences. Researchers who are used to working with GLIM may prefer the HT version despite its two step nature, whereas those familiar with Epicure may prefer the CT version.

Also Oksanen (in an unpublished doctoral dissertation from 1998) has modelled “normal” mortality, rather than treating it as a constant.<sup>15</sup> Only age and period were included in this first additive term but, unless computational problems are met, it should be possible to include also other variables. The data source for “normal” mortality was national aggregate tables on deaths and size of subpopulations in different cells, from which the corresponding numbers for cancer patients were subtracted. After this data manipulation, estimation was carried out in GLIM.

### Comparison with model 1b

The differences between estimates from models 1a and 1b can be easily understood. The idea behind model 1a is that survival

## Key points

- The model (1a) proposed here permits a broad range of sociodemographic variations in “normal” mortality to be taken into account, and is a convenient alternative to an existing technique.
- If sufficient data on sociodemographic variations in “normal” population are not available, the best solution is to model registered cancer mortality among cancer patients (model 2b).
- If cause specific mortality data are not available either, there is little need to bother with the estimation of a relative survival model (1b). One may just as well estimate all cause mortality among cancer patients (model 2a).
- Even the least data demanding models (1b and 2a) perform well for the most aggressive cancers.

is given by the mortality among cancer patients minus a “normal” mortality that may vary with sociodemographic characteristics. In model 1b, however, it is a constant “normal” mortality that is subtracted. For example, the true excess mortality ( $\mu_{c,*}$ ) at educational level  $E^*$  is lower than found when using a constant ‘normal’ mortality ( $\mu_{c,*}$ ), and that at a higher educational level  $E^{**}$  is higher (fig 1). Stated differently, the higher mortality observed for cancer patients with education  $E^*$  according to model 1b is partly attributable to the fact that they would have had a higher mortality than those with  $E^{**}$  also in the absence of the disease.

As expected, the bias is especially large when groups with widely different “normal” mortality are studied. For example, the excess mortality of divorced men (compared with that of the married) is particularly high according to model 1b.

Although the differentials in the “normal” mortality are substantial on a relative scale (up to 79%), failure to correct for them does not matter much when this “normal” mortality is generally low compared with cancer patients’ mortality. This is the reason why estimates from model 1a and 1b are so similar for lung cancer, pancreas cancer, and stomach cancer, whereas they are more different for prostate cancer, bladder cancer, and malignant melanoma.

For the same reasons, differences were also found to be extra large when only the localised cancers were considered (not shown). For example, when all localised cancers in men were pooled together, the point estimates for marital status effects differed by as much as 0.07–0.15 between models 1a and 1b. For localised breast cancer in women, the corresponding figures were 0.08–0.14, which is considerably larger than the differences of 0.03–0.05 estimated in models for all breast cancers. Conversely, when a model was estimated only for prostate cancer with distant metastases, differences of 0.04–0.09 were found, whereas they were 0.09–0.22 for all prostate cancers.

If follow up had been increased to, say, 10 years, mortality among cancer patients would have been more dominated by “normal” mortality, and one would expect social gradients in survival according to model 1b to differ even more from those obtained with model 1a. This has not been checked here, but the Finnish study<sup>9</sup> suggests such a pattern.

### Comparison with models 2a and 2b

Elimination of the first term (which corresponds to subtracting 0) gives many estimates closer to 1, and thus also closer to those from model 1a. This can be intuitively understood by taking a new look at figure 1. Given age, sex, and period, the excess mortality at educational level  $E^*$  according to model 1b is  $\mu_{c,*}$  and that at educational level  $E^{**}$  is  $\mu_{c,**}$ . When no correction at all is made for “normal” mortality, the ratio between the two levels would be  $(\mu_{c,*} + \mu_{n,*})/(\mu_{c,**} + \mu_{n,**})$ , which is larger than  $\mu_{c,*}/\mu_{c,**}$  if the latter is  $< 1$  and smaller if it is  $> 1$ . However, in a sample consisting of several ages and periods (as in this sex specific analysis), estimates will not always be

closer to 1. An estimate more different from 1 is seen, for example, for cervical cancer for better-educated women. (One particularly large difference is found for the few divorced women with malignant melanoma.)

Social and marital differentials in non-cancer deaths are usually found to be sharper than those in cancer deaths.<sup>16, 17</sup> This is consistent with the fact that exclusion of non-cancer deaths yields less pronounced effects of education and marital status (according to a comparison of results from model 2b with those from 2a).

As Auvinen *et al*<sup>18</sup> rightly pointed out, the excess mortality caused by a malignancy is higher than the registered cancer mortality among cancer patients and lower than their all cause mortality. For that reason, they showed estimates from both kind of models in their study of social inequalities in cancer survival. However, it cannot be concluded that the social or marital differentials in the true excess mortality lie between those in registered cancer mortality and those in all cause mortality.

On the contrary, this study has shown that both the results from model 2a and those from 2b tend to be overestimates of the differentials in the true aggressiveness of the disease (considered to be captured by model 1a). However, the differences are not very large, which should be a great comfort to researchers without access to data on sociodemographic variations in general mortality. For many important cancer sites, the differentials in prognosis between those who are married and those who are not, and between the poorly educated and the better educated, can be analysed from a patient material exclusively, above all if data permit non-cancer deaths to be excluded. The differences between models 2b and 1a for all cancers pooled were only 0.00–0.02. For some cancer sites, however, the choice of method is more important, although it can hardly be said to be critical. For example, while never married men were found to have an excess mortality from prostate cancer that was 14% higher than that of the married, according to the most complex model (1a), the corresponding parameter was 20% in the cancer mortality model estimated for cancer patients (2b) (and it was 24% when only an age specific and period specific “normal” mortality was used in the relative survival approach in model 1b). Similarly, a model for registered prostate cancer mortality gave an almost significant excess mortality of 6% for the widowed, while the estimate from the most complex model was 0. (The other models gave significant effects of 6% or 7%).

### The bottom line

When the intention is to assess, for example, social or marital variations in survival within a relative survival approach, corresponding variations in “normal” mortality should be taken into account if possible. (Such variations are, of course, of no concern when the focus is on other determinants of cancer survival). In accordance with results reported from Finland, effects can be markedly different when “normal” mortality is assumed to depend only on age, sex, and period. The most extreme example is the doubling of the marital status differentials in survival from prostate cancer. Differences are smaller for the most aggressive cancers that strongly dominate mortality. Because the use of the simplest relative survival model exaggerates the steepness of the survival gradients for the less aggressive cancers in particular, and because these cancers also tend to respond better to treatment than others, results from this model may be taken as more supportive of the treatment explanation (compared with the host factor explanation; see arguments in introduction) than they should be.

Model 1a allows controls for sociodemographic variations (beyond those related to age, sex, and period) in the population without the cancer diagnosis under investigation. It is also practically convenient, but requires individual data including both persons with and persons without cancer (although not necessarily the whole national population). Also the Hakulinen-Tenkanen version can, of course, be used when the data required

for model 1a are available. This approach permits similar controls for sociodemographic variations in “normal” mortality, but is in some respects less satisfying. One should expect the results to be very similar, but an empirical check of that is left to future studies. On the other hand, the Hakulinen-Tenkanen version has the advantage that it can be used also when only aggregate data are available for the “normal” population.

The results also confirm the conclusion from the Finnish study that, when data do not permit a sufficiently broad range of sociodemographic variations in “normal” mortality to be taken into account, the best alternative is to model registered cancer mortality among cancer patients (model 2b). This model gives results that are somewhat different from those from model 1a, but the differences can hardly be said to be large. If data on cause of death are not available either, one may well choose the simplest model (2a). This model was not estimated in the Finnish study, but it is found here that it performs better than the technically more cumbersome relative survival approach (1b), using the most complex model as the yardstick. Fortunately, both these less data demanding models compare well with the most complex model for the most aggressive cancers.

### ACKNOWLEDGEMENTS

Thanks are due to Statistics Norway for allowing the use of the data. Comments from Dr Steinar Tretli and two anonymous referees are greatly appreciated.

Conflicts of interest: none.

### REFERENCES

- 1 Kravdal Ø. Social inequalities in cancer survival. *Population Studies* 2000;**54**:1–18.
- 2 Kravdal Ø. The impact of marital status on cancer survival. *Soc Sci Med* 2001;**52**:357–68.
- 3 Krongrad A, Lai H, Burke MA, *et al*. Marriage and mortality in prostate cancer. *J Urol* 1996;**156**:1696–700.
- 4 Neale AV. Racial and marital status influences on 10 year survival from breast cancer. *J Clin Epidemiol* 1994;**47**:475–83.
- 5 Gilliland FD, Hunt WC, Morris DM, *et al*. Prognostic factors for thyroid carcinoma. A population-based study of 15,698 cases from the Surveillance, Epidemiology and End Results (SEER) program 1973–1991. *Cancer* 1997;**79**:564–73.
- 6 Schrijvers CTM, Coebergh JWW, van der Heiden LH, *et al*. Socioeconomic variation in cancer survival in the southeastern Netherlands, 1980–89. *Cancer* 1995;**75**:2946–53.
- 7 Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society* 1987;**36**:309–17.
- 8 Esteve J, Benhamou JE, Croasdale M, *et al*. Relative survival and the estimation of net survival: elements for further discussion. *Stat Med* 1990;**9**:529–38.
- 9 Dickman PW, Auvinen A, Voutilainen ET, *et al*. Measuring social class differences in cancer patient survival: Is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *J Epidemiol Community Health* 1998;**52**:727–34.
- 10 Harvei S, Kravdal Ø. The importance of marital and socioeconomic status in incidence and survival of prostate cancer. *Prev Med* 1997;**26**:623–32.
- 11 Kravdal Ø, Hansen S. The importance of childbearing for Hodgkin's disease: new evidence from incidence and mortality models. *Int J Epidemiol* 1996;**25**:737–43.
- 12 Agresti A. *Categorical data analysis*. New York: Wiley, 1990.
- 13 Kravdal Ø. Is the relationship between childbearing and cancer incidence due to biology or lifestyle? Examples of the importance of using data on men. *Int J Epidemiol* 1995;**24**:477–84.
- 14 Preston DL, Rubin JH, Pierce PA, *et al*. *Epicure user's guide*. Seattle: Hirosoft International Corporation, 1993.
- 15 Oksanen H. *Modelling of survival of prostate cancer patients*. Tampere: Acta Universitatis Tampensis 615, University of Tampere, Finland, 1998.
- 16 Jung IM, Glerum JJ, Van Poppel FW, *et al*. The contribution of specific causes of death to mortality differences by marital status in the Netherlands. *Eur J Public Health* 1996;**6**:142–9.
- 17 Vägerö D, Lundberg O. Socioeconomic mortality differentials among adults in Sweden. In: Lopez A, Caselli G, Valkonen T, eds. *Adult mortality in developed countries: from description to explanation*. Oxford: Clarendon Press, 1995:223–42.
- 18 Auvinen A, Karjalainen S, Pukkala E. Social class and cancer patient survival in Finland. *Am J Epidemiol* 1995;**142**:1089–102.