Regression models for relative survival

Paul W. Dickman[1,*,†], Andy Sloggett[2], Michael Hills[3] and Timo Hakulinen[4]

[1] *Department of Medical Epidemiology, Karolinska Institutet, Stockholm, Sweden*
[2] *London School of Hygiene & Tropical Medicine, London, U.K.*
[3] *Retired*
[4] *Finnish Cancer Registry, Helsinki, Finland and Department of Public Health, University of Helsinki, Finland*

SUMMARY

Four approaches to estimating a regression model for relative survival using the method of maximum likelihood are described and compared. The underlying model is an additive hazards model where the total hazard is written as the sum of the known baseline hazard and the excess hazard associated with a diagnosis of cancer. The excess hazards are assumed to be constant within pre-specified bands of follow-up. The likelihood can be maximized directly or in the framework of generalized linear models. Minor differences exist due to, for example, the way the data are presented (individual, aggregated or grouped), and in some assumptions (e.g. distributional assumptions). The four approaches are applied to two real data sets and produce very similar estimates even when the assumption of proportional excess hazards is violated. The choice of approach to use in practice can, therefore, be guided by ease of use and availability of software. We recommend using a generalized linear model with a Poisson error structure based on collapsed data using exact survival times. The model can be estimated in any software package that estimates GLMs with user-defined link functions (including SAS, Stata, S-plus, and R) and utilizes the theory of generalized linear models for assessing goodness-of-fit and studying regression diagnostics. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: relative survival; excess mortality; net survival; cancer; registry; regression

1. INTRODUCTION

A common aim when studying cancer patient survival is the estimation of net survival, a measure of patient survival corrected for the effect of other causes of death. Net survival is a hypothetical quantity which can be estimated using, for example, cause-specific survival or relative survival. In studies of cancer patient survival conducted in a clinical setting, whether they be randomized trials [1] or observational studies [2], it is standard to use cause-specific survival to estimate net survival. Estimates of cause-specific survival are obtained by considering the survival times of patients who died of causes other than the cancer of interest

to be censored and applying the usual actuarial or product-limit estimator. The definition of net survival is not consistent in the statistical and epidemiological literature. Some authors, for example, use the term net survival as a synonym for cause-specific survival. Our use of the term is consistent with the use of the term 'net probability' in the theory of competing risks [3, 4].

Our focus is the estimation and modelling of net survival in population-based studies of cancer survival, that is, studies based on data collected by population-based cancer registries. Although the basic methodology is similar to that used in clinical studies of cancer patient survival the two types of studies are, in many ways, quite different. The aim of a randomized clinical trial is usually to evaluate the effect of an intervention (including treatment) on patient survival while controlling for other factors which may affect patient survival. It is not essential that the patients under study be representative of any population or population group, although it is essential that the treatment groups being compared are comparable in all aspects other than the intervention, which is best achieved by randomization. The aim of population-based studies of cancer survival, on the other hand, is to describe patient survival in demographically defined groups in the population in such a way that the results are representative of the population. The population-based design is vital from a public health perspective, but such studies are also useful from a clinical perspective. However, a randomized clinical trial is clearly preferable if the primary interest is assessing treatment efficacy since it is essentially impossible to control for all of the factors associated with treatment allocation in an observational setting.

In general, population-based studies involve a larger number of patients who are followed for a longer period, although less accurately with respect to clinical outcomes such as cause of death, relapses, remissions, side effects, etc. As such, use of cause-specific survival methods can be problematic since information on cause of death is often unavailable or unreliable [5]. It can happen, for example, that when a colon cancer metastasises to the liver and causes death that the cause of death on the death certificate is erroneously recorded as cancer of the liver. In a cause-specific survival analysis of patients diagnosed with colon cancer, this will be classified as a 'death due to other causes' and the survival time will be considered censored at the time of death. Even if accurate information on cause of death is available, it is often difficult to determine whether or not a death should be classified as being due to the cancer of interest. For example, it is not obvious how one should classify deaths due to suicide or the secondary effects of treatment. It is not possible in cause-specific survival analysis to classify a death as being partially due to cancer; the only two alternatives are to classify a death as being entirely due to the cancer of interest or entirely due to other causes.

Because of these difficulties, it is common to use relative survival as a means of estimating net survival in population-based cancer survival studies. Relative survival is generally estimated from life tables as the ratio of the observed survival of the patients (where all deaths are considered events) to the expected survival of a comparable group from the general population, matched to the patients with respect to the main factors affecting patient survival and assumed to be practically free of the cancer of interest [6]. It is usual to estimate expected survival from nationwide population life tables stratified by age, sex, calendar time and where applicable, race [7]. Although these tables include the effect of deaths due to the cancer being studied, this does not, in practice, affect the estimated survival proportions [6,8, pp. 235–236]. Mortality for a specific site generally constitutes only a small fraction of total mortality and correcting for this mortality has a negligible effect on estimates of expected survival, even among common cancers such as prostate cancer [9]. The major advantages of relative

survival are that information on cause of death is not required and that it provides a measure of the excess mortality experienced by patients diagnosed with cancer, irrespective of whether the excess mortality is directly or indirectly attributable to the cancer. Relative survival is now accepted as the method of choice for population-based cancer survival. Population-based cancer registries almost exclusively report relative survival as the measure of patient survival, prominent recent examples being reports from the U.S.A. [10], England and Wales [11] and the EUROCARE-2 [12] study of 17 European countries.

When modelling relative survival, the hazard function at time t since diagnosis for persons diagnosed with cancer (with covariate vector $\mathbf{z}$) is modelled as the sum of the expected hazard, $\lambda^*(t;\mathbf{z})$, and the excess hazard due to a diagnosis of cancer, $v(t;\mathbf{z})$. That is

$$\lambda(t;\mathbf{z}) = \lambda^*(t;\mathbf{z}) + v(t;\mathbf{z}) \tag{1}$$

The expected hazard (sometimes called the baseline hazard) is denoted with an asterisk to indicate that it is estimated from external data (general-population mortality rates) as opposed to, for example, the baseline hazard in a Cox proportional hazards model [13], an arbitrary function which is not estimated. Some authors prefer to write the expected hazard as $\lambda^*(t;\mathbf{z}_1)$, where $\mathbf{z}_1$ is a subvector of $\mathbf{z}$, in order to indicate that the expected hazard is generally assumed to depend only on a subset of the covariates available (typically age, sex and period). The expected hazard does not depend, for example, on tumour-specific covariates such as histology or stage. We will write, for simplicity, that the expected hazard is a function of $\mathbf{z}$, even though it does not vary over all elements of $\mathbf{z}$.

The model is known as an additive hazards model or a relative survival model, since it can be written as

$$S(t;\mathbf{z}) = S^*(t;\mathbf{z}) \times r(t;\mathbf{z}) \tag{2}$$

where $S(t;\mathbf{z})$, $S^*(t;\mathbf{z})$ and $r(t;\mathbf{z})$ represent cumulative observed, expected and relative survival. For population-based cancer survival data, such models are generally biologically more plausible and provide a better fit to the data than multiplicative models [14–17]. The hazards are assumed to be constant within pre-specified subintervals (bands) of follow-up time (i.e. piecewise constant hazards). These intervals are typically of length 1 year, although it is common to use shorter intervals early in the follow up (e.g. during the first year) and longer intervals later in the follow-up (e.g. after 10 years). A set of indicator variables is constructed (one indicator variable for each interval excluding the reference interval) and incorporated into the covariate vector. We will use $\mathbf{x}$ to denote the covariate vector that contains indicator variables for these bands of follow-up time in addition to the other covariates $\mathbf{z}$. Our primary interest is in the excess hazard component, v, which is assumed to be a multiplicative function of the covariates, written as $\exp(\mathbf{x}\beta)$. The basic relative survival model is therefore written as

$$\lambda(\mathbf{x}) = \lambda^*(\mathbf{x}) + \exp(\mathbf{x}\beta) \tag{3}$$

Parameters representing the effect in each follow-up interval are estimated in the same way as parameters representing the effect of, for example, age, sex or histology. Implicit in equation (3) is the assumption that the excess hazards for any two patient subgroups are proportional over follow-up time. Non-proportional excess hazards can, however, be incorporated by including time by covariate interaction terms in the model.

The focus of this paper is on methods for estimating the model based on patient data and a file of general-population mortality rates (or probabilities of death). The exponentiated parameter estimates have an interpretation as excess hazard ratios, sometimes known as relative excess risks [18]. An excess hazard ratio of, for example, 1.5 for males compared to females implies that the excess hazard associated with a diagnosis of cancer is 50 per cent, higher for males than females.

Two approaches to estimating the model have previously been described by Hakulinen and Tenkanen [15] and Estève *et al.* [16]. Hakulinen and Tenkanen estimate the model based on grouped data in the framework of generalized linear models (GLMs) using a binomial assumption for the number of observed deaths. Estève *et al.* use a full-likelihood approach based on individual level data (exact survival times). We will show, by using an approach analogous to that used in the analysis of epidemiological cohort studies, that the likelihood can be written in a way simpler than that used by Estève *et al.* thereby enabling estimation using a full-likelihood approach with standard statistical software. We will also describe a new approach whereby the model is estimated in the framework of generalized linear models using a Poisson assumption for the number of observed deaths. This approach can be applied to either aggregated data or individual level data (exact survival times). When applied to individual level data using information on exact survival times, the estimates from the Poisson GLM are identical to those obtained using the full-likelihood approach. We apply the four approaches to estimating the model to two data sets kindly provided by the Finnish Cancer Registry.

2. THE ESTÈVE *ET AL.* FULL-LIKELIHOOD APPROACH

Estève *et al.* [16] described a method for estimating the model in equation (3) directly from individual-level data using a full-maximum-likelihood approach. Although they use a slightly different parameterization in their paper, the underlying model is identical to equation (3).

The likelihood function is

$$L = \prod_{i=1}^{n} \exp\left(-\int_{0}^{t_i} \lambda(s)\,ds\right) [\lambda(t_i)]^{d_i} \qquad (4)$$

where t_i is the survival time and d_i the failure indicator variable (1 if t_i is the time of death; 0 if the survival time is censored at t_i) for each of the $i = 1, \ldots, n$ individuals.

Writing the total hazard as the sum of the expected hazard and the excess hazard, the log-likelihood function is

$$l(\beta) = -\sum_{i=1}^{n} \int_{0}^{t_i} \lambda^*(s)\,ds - \sum_{i=1}^{n} \int_{0}^{t_i} \nu(s)\,ds + \sum_{i=1}^{n} d_i \ln[\lambda^*(t_i) + \nu(t_i)] \qquad (5)$$

Although the model is specified in continuous time, it is assumed, as with all approaches described in this paper, that the hazard is constant within pre-specified bands of time and the excess hazard $\nu(t)$ is written as $\exp(\mathbf{x}\beta)$. The first component of the log likelihood does not depend on β leading to the attractive feature from a computational viewpoint that, for each individual, only one value needs to be read from the excess hazards file, the expected hazard at t_i.

Their approach is implemented in special-purpose software which runs under DOS [19]. A major problem in applying the Estève *et al.* approach in practice is that it is not possible, using the accompanying software, to model time-varying covariates. This means that there is

no way to control for non-proportional excess hazards, which are very common with cancer registry data. It should be noted that this problem is due to limitations in the associated software rather than the theory underlying the model. Regression diagnostics are not available for the Estève *et al.* model and there is no way of assessing goodness-of-fit (due to the lack of underlying theory rather than limitations in the associated software).

3. FULL-LIKELIHOOD BASED ON MULTIPLE OBSERVATIONS PER SUBJECT

In the Estève *et al.* approach, as with all approaches described in this paper, the excess hazard is assumed to be constant within bands of follow-up time. Estimation of the model is simplified if each observation is split into separate observations for each band of follow-up. Rather than evaluating the log-likelihood for each subject and summing over subjects (the Estève *et al.* approach) we evaluate the log-likelihood for each subject band. Consider, for example, an individual who dies 4.5 years after diagnosis ($t_i = 4.5$, $d_i = 1$). This observation is split into five subject-band observations, for which the time at risk is $y = 1$ year and the censoring indicator $d = 0$ for the first four whereas $y = 0.5$ and $d = 1$ for the fifth subject-band observation.

Each of the subject-band observations inherits the covariates of the original observation (age at diagnosis, sex, stage, etc.). Estimates of $\lambda^*(\mathbf{x})$ for each subject band are made using external data (general-population mortality rates). Each subject-band observation, indexed by j, represents the survival experience of an individual patient during a pre-specified band of follow-up and includes variables representing the time at risk (y_j), death indicator (d_j), expected hazard (λ_j^*) and indicator variables for each of the components of β (including follow-up band).

The log-likelihood function, expressed in terms of the J subject-band observations, is

$$l(\beta) = \sum_{j=1}^{J} [d_j \ln[\lambda^*(\mathbf{x}_j) + \exp(\mathbf{x}_j\beta)] - y_j \exp(\mathbf{x}_j\beta)] \qquad (6)$$

This log-likelihood is obtained in an identical fashion to equation (5). The first component in equation (5) does not depend on β and so can be omitted. Since the excess hazard v is assumed to be constant in each interval, the integral in the second component of equation (5) evaluates to the interval length (y_j) multiplied by the excess hazard (written as $\exp(\mathbf{x}_j\beta)$). The model can be estimated using procedures available in standard statistical software packages for maximum-likelihood estimation, such as the Stata `ml` command or SAS PROC NLP (part of SAS/OR). It is a simple matter to model non-proportional excess hazards by estimating appropriate time by covariate interaction terms. Regression diagnostics are, however, not available and there is no way of assessing goodness-of-fit (due to the lack of underlying theory).

4. A GENERALIZED LINEAR MODEL WHERE THE OBSERVED NUMBER OF DEATHS IS ASSUMED POISSON

The relative survival model (equation (3)) assumes piecewise constant hazards which implies a Poisson process for the number of deaths in each interval; see Reference [20, p. 409] or Reference [21, Section 4.2]. The resulting log likelihood (equation (6)) is identical to the

log-likelihood for grouped Poisson data with intensity $\lambda^*(\mathbf{x}) + \exp(\mathbf{x}\beta)$ [21, p. 185] except we omitted the term $-y_j\lambda^*(\mathbf{x}_j)$ since it did not depend on β.

This implies that the relative survival model can also be estimated in the framework of generalized linear models using a Poisson assumption for the observed number of deaths. If the model is estimated from subject-band observations, the estimates will be identical to those obtained using the full-likelihood approach (Section 3) since we maximize the same likelihood based on the same data. We can also, however, estimate the model based on collapsed or grouped data, in which case the estimates differ slightly.

We assume that the number of deaths, d_j, for observation j can be described by a Poisson distribution, $d_j \sim \text{Poisson}(\mu_j)$ where $\mu_j = \lambda_j y_j$ and y_j is person-time at risk for the observation. The observations can represent either life table intervals (in which case there can be multiple deaths per observation), individual patients or subject bands (as in Section 3).

Equation (3) is then written as

$$\mu_j/y_j = d_j^*/y_j + \exp(\mathbf{x}\beta) \tag{7}$$

which can be written as

$$\ln(\mu_j - d_j^*) = \ln(y_j) + \mathbf{x}\beta \tag{8}$$

where d_j^* is the expected number of deaths (due to causes other than the cancer of interest and estimated from general population mortality rates). This implies a generalized linear model with outcome d_j, Poisson error structure, link $\ln(\mu_j - d_j^*)$, and offset $\ln(y_j)$. Breslow and Day [21, pp. 173–176] discuss similar models with application to an occupational cohort study. Because of the non-standard link function, fitting the model requires software which supports the estimation of generalized linear models with the so-called user-defined link functions. Most general purpose statistical software packages support this feature, including SAS (from version 6.10), Stata (from version 7), S-plus, R and GLIM.

Since the data are usually cross-classified and non-sparse, evaluation of model goodness-of-fit can be made using the deviance or Pearson χ^2 statistics [22] although most software packages require the data to be collapsed such that there is only one observation for each unique combination of covariates. The usual regression diagnostics (residuals, influence statistics) applicable for generalized linear models are also available. It should be noted that we introduced the assumption that $d_j \sim \text{Poisson}(\mu_j)$ in order to use the GLM approach, although this is not strictly necessary—we simply make use of the fact that the likelihood in equation (6) is identical to a Poisson likelihood.

4.1. Estimation based on collapsed data (using exact survival times)

The model can be estimated directly from subject-band observations or the subject-band observations can be collapsed to give one observation for each covariate pattern (d, d^* and y are summed within each covariate pattern). Estimating a standard Poisson regression model (with logarithmic link and offset $\ln(y_j)$) gives identical estimates for both individual and collapsed data. Estimating equation (8) based on collapsed data, however, leads to slightly different estimates to those obtained from subject-band observations since d^* varies within each covariate pattern (i.e. combination of follow-up interval, sex, period, age group, etc.).

4.2. Estimation based on grouped data

Grouped survival data occur when information is available only on the number of events in a time interval rather than exact times to event for each individual. Grouped survival data arise when, for example, patients are contacted at fixed intervals after the date of diagnosis (e.g. annually) in order to ascertain vital status or when more accurate information is available on survival times but the data are grouped using actuarial methods. Many methods for the analysis of time-to-event data assume that survival time is measured on a continuous scale although, in practice, survival time is always measured on a discrete scale.

Survival time is calculated from the date of diagnosis of cancer to the date of death. The date of diagnosis is not a well-defined quantity since diagnosis of cancer is generally a process which commences with a suspicion and becomes more definitive through X-rays, endoscopies and finally, microscopic investigation [23]. For this reason many registries, including the Finnish Cancer Registry who provided the data studied in this paper, record survival time in completed months rather than completed days. For cancer registry data, it is usual to assume that survival times recorded in completed days or completed months may be analysed using methods developed for continuous time, but discrete time methods should be used when the data are more heavily grouped.

In the analysis of cancer patient survival, it is common to estimate relative survival using actuarial (life-table) methods. The patients are grouped into K strata, indexed by k, with one stratum for each combination of relevant predictor variables (age, sex, calendar period of diagnosis, stage, etc.), and a life-table, with intervals indexed by i, estimated for each stratum.

We will use the following notation for quantities derived from life tables:

n_{ki} number of individuals alive at the start of the ith life-table interval,

d_{ki} number of deaths during the ith interval,

w_{ki} number censored during the ith interval,

l'_{ki} effective number at risk ($l'_{ki} = n_{ki} - w_{ki}/2$),

y_{ki} total person-time at risk during the ith interval,

p^*_{ki} estimated interval-specific expected survival proportion (estimated from general population mortality rates),

d^*_{ki} expected number of deaths (due to causes other than cancer and estimated from general population mortality rates) for the ith interval,

λ^*_{ki} expected hazard (due to causes other than cancer and estimated from general population mortality rates) for the ith interval.

When estimating the relative survival model (equation (3)) from grouped data, the observations are the life-table intervals. Ideally, person-time at risk should be estimated from information on exact survival times if they are available. If, however, only grouped survival data are available then person-time at risk must be approximated. If grouped survival data are available based on annual life-table intervals, then an approximation for the number of person-years at risk during an interval is $y_{ki} = n_{ki} - (w_{ki} + d_{ki})/2$. This approximation implies an assumption that deaths and censorings are evenly distributed throughout an interval, an assumption which is generally valid except sometimes for the first interval, where a correction factor can be applied, if necessary, or shorter intervals can be used at the start of follow-up.

The expected number of deaths during an interval can be approximated by either

$$d_{ki}^* = (n_{ki} - w_{ki}/2)(1 - p_{ki}^*) \tag{9}$$

or

$$d_{ki}^* = -\ln(p_{ki}^*/\Delta_{ki})y_{ki} \tag{10}$$

where Δ_{ki} is the length of the interval, depending on whether we choose to work with proportions (equation (9)) or rates (equation (10)).

The distinction between what we have called collapsed data and grouped data is that, in the collapsed data, y and d^* are based on the exact time at risk whereas these quantities are approximated for grouped data. The collapsed data set will contain the same number of observations as the grouped data set with identical values of d in each. The primary reason for collapsing the data is that residuals and goodness-of-fit statistics are not appropriate when estimated from subject-band (or individual level) observations.

5. THE HAKULINEN–TENKANEN APPROACH FOR GROUPED DATA

Hakulinen and Tenkanen [15] estimate the relative survival model from grouped survival data using an assumption that the number of deaths in each life-table interval can be modelled using a binomial distribution. The model is estimated in the framework of generalized linear models [22] where the outcome is $l_{ki}' - d_{ki}$ (the number of patients surviving the interval), the error structure binomial with denominator l_{ki}' and the link function complementary log–log combined with a division by p_{ki}^*. That is

$$\ln\left[-\ln\frac{p_{ki}}{p_{ki}^*}\right] = \mathbf{x}\beta \tag{11}$$

As with the approaches described in Section 4, the model can be estimated using any software which supports the estimation of generalized linear models with user-defined link functions. The binomial assumption has been criticized [16] since the patients at risk at the start of each interval are known to be heterogeneous so will not have the same probability of surviving to the end of the interval. This problem can be minimized, however, by ensuring that the life-table estimates are stratified by, for example, age and stage. If considered necessary, adjustment can be made for extra-binomial variation by scaling the covariance matrix. The two approaches based on grouped data will give very similar results since the log-likelihoods are similar. The binomial log-likelihood for a life-table interval [24, p. 23] is

$$d_{ki}\log(1 - \mathrm{e}^{-\lambda_{ki}}) + (l_{ki}' - d_{ki})\log(\mathrm{e}^{-\lambda_{ki}}) \tag{12}$$

which, for small λ_{ki}, can be written as

$$d_{ki}\log(\lambda_{ki}) - (l_{ki}' - d_{ki})\lambda_{ki} \tag{13}$$

The Poisson log-likelihood, where person-time is approximated as $y_{ki} = l_{ki}' - 0.5d_{ki}$, is

$$d_{ki}\log(\lambda_{ki}) - (l_{ki}' - 0.5d_{ki})\lambda_{ki} \tag{14}$$

which will be similar to the binomial log-likelihood when λ_{ki} is small.

6. EMPIRICAL COMPARISON OF THE MODELS

In this section, four approaches to estimating the relative survival model (equation (3)) are applied to two data sets kindly supplied by the Finnish Cancer Registry. The first data set consists of 5318 patients diagnosed with localized skin melanoma and the second consists of 6274 patients diagnosed with localized colon carcinoma. The data sets contain all cases diagnosed in Finland (population 5.1 million) during 1975–1994 with follow-up to the end of 1995 [25]. The localized skin melanoma data were chosen since this was one of the few cancer sites for which a main-effects model provides a reasonable fit to the data (i.e. an assumption of proportional excess hazards is appropriate). In general, excess hazards are almost always non-proportional with respect to stage and, even within each stratum of stage, excess hazards are usually non-proportional with respect to age. The localized colon carcinoma data provide a typical example of data exhibiting non-proportional excess hazards with respect to age, an issue which is discussed further in Section 6.1.

The following approaches to estimating the model were used:

(1) Grouped survival times, GLM with a binomial error structure (Section 5);
(2) grouped survival times, GLM with a Poisson error structure (Section 4.2);
(3) exact survival times, individual subject-band observations (estimates are identical using either the full-likelihood approach (Section 3) or a GLM with a Poisson error structure (Section 4)); and
(4) exact survival times, collapsed data, GLM with a Poisson error structure (Section 4.1).

The models were estimated for the first 5 years of follow-up only. That is, patients alive on the fifth anniversary following diagnosis were considered censored on that date. This was done since the assumption of proportional excess hazards is not generally appropriate for longer follow-up periods and the majority of the excess deaths occur in the first 5 years following diagnosis. Follow-up time was stratified into annual intervals. Estimates are presented in Table I as estimated excess hazard ratios compared to the appropriate reference category (first year of follow-up, males, period of diagnosis 1975–1984, ages 0–44 years at diagnosis).

One can see from the estimated excess hazard ratios in Table I, for example, that excess mortality following a diagnosis of cancer has decreased with calendar period of diagnosis (survival has improved) for both cohorts. The estimated relative excess risk of 0.63 for calendar period indicates that patients diagnosed with localized skin melanoma during 1985–1994 experienced only 63 per cent of the excess mortality experienced by those diagnosed 1975–1984. The difference is statistically significant (a Wald test statistic is $\ln(0.63)/0.098 = -4.7$). In practice, tests of statistical significance are best performed using the likelihood ratio criterion. It is not surprising that the estimated parameters and standard errors presented in Table I are similar for each of the four approaches, since each estimates the same underlying model (equation (3)). All approaches assume that the excess hazard is constant within each follow-up band and all require the widths of the bands to be pre-specified. As expected, the estimates are slightly different when grouped, rather than exact, survival times are used (models (1) and (2)). The primary difference between the two models for grouped survival data is the assumed error distribution (binomial or Poisson).

Since approaches (1), (2) and (4) estimate the model based on non-sparse cross-classified data in the framework of generalized linear models it is possible to use the deviance statistic as a measure of goodness-of-fit. Under the assumption that the model provides an adequate

Table I. Estimated excess hazard ratios and standard errors of the log excess hazard ratio for four approaches to estimating the relative survival model for two data sets, localized skin melanoma and localized colon carcinoma diagnosed in Finland during 1975–1994 with follow-up to the end of 1995.

	Skin melanoma				Colon carcinoma			
	Grouped		Exact times		Grouped		Exact times	
	Bin. (1)	Pol. (2)	Ind. (3)	Coll. (4)	Bin. (1)	Pol. (2)	Ind. (3)	Coll. (4)
Deviance	76	73		76	120	113		131
Residual df	70	70		70	70	70		70
Estimated excess hazard ratios (*i.e.* exp (β)))								
Follow-up 2/1	6.69	6.64	6.79	6.76	0.84	0.85	0.83	0.80
Follow-up 3/1	7.11	7.07	7.13	7.24	0.65	0.66	0.68	0.62
Follow-up 4/1	5.33	5.30	5.36	5.42	0.52	0.53	0.54	0.50
Follow-up 5/1	4.59	4.56	4.73	4.66	0.45	0.46	0.46	0.43
Female/Male	0.56	0.57	0.55	0.56	0.96	0.98	0.95	0.96
Year 85−94/75−84	0.63	0.63	0.63	0.63	0.73	0.73	0.73	0.73
Age 45−59/0−44	1.38	1.38	1.38	1.38	0.86	0.86	0.87	0.86
Age 60−74/0−44	1.90	1.86	1.92	1.89	1.07	1.05	1.06	1.07
Age 75 + /0−44	3.19	2.99	3.14	3.24	1.37	1.29	1.34	1.44
SEs of the log hazard ratio (*i.e. of the parameter estimates*)								
Follow-up 2/1	0.298	0.301	0.297	0.301	0.093	0.095	0.094	0.092
Follow-up 3/1	0.299	0.301	0.298	0.301	0.109	0.111	0.108	0.108
Follow-up 4/1	0.307	0.310	0.306	0.309	0.131	0.133	0.128	0.130
Follow-up 5/1	0.315	0.317	0.313	0.317	0.151	0.153	0.150	0.150
Female/Male	0.097	0.098	0.097	0.097	0.077	0.079	0.077	0.076
Year 85−94/75−84	0.098	0.099	0.097	0.098	0.075	0.076	0.075	0.074
Age 45−59/0−44	0.125	0.125	0.125	0.125	0.156	0.157	0.156	0.157
Age 60−74/0−44	0.128	0.129	0.127	0.128	0.143	0.144	0.143	0.143
Age 75 + /0−44	0.173	0.181	0.173	0.172	0.151	0.153	0.151	0.150

fit to the data, the deviance will follow a χ^2 distribution with degrees of freedom equal to the number of residual degrees of freedom of the model (number of observations minus number of estimated parameters) [22].

There is no evidence of lack-of-fit for the model fitted to the skin melanoma data since the deviance is similar in magnitude to the residual degrees of freedom. A complete assessment of model fit should also include a study of residuals and influence statistics, which are not shown here although they show no evidence of lack-of-fit. There is, however, clear evidence of a lack-of-fit for the model fitted to the localized colon data. The deviance is 113 on 70 residual degrees of freedom (model (2) in Table I) whereas the 99th percentile of a χ^2_{70} distribution is 100. This issue is discussed further in the following section. It is worth noting, however, that all approaches to fitting the model produce similar estimates even when the model is poorly specified.

Table II. Estimated excess hazard ratios for age estimated separately for each annual follow-up interval by including an age by follow-up interaction in model (5), localized colon carcinoma diagnosed in Finland during 1975–1994 with follow-up to the end of 1995.

Age	Follow-up interval				
	1	2	3	4	5
0–44	1.00	1.00	1.00	1.00	1.00
45–59	1.10	0.59	1.22	0.72	0.97
60–74	1.66	0.89	1.02	0.85	0.97
75+	3.31	0.83	0.63	0.52	0.01

6.1. Non-proportional excess hazards

For a full main effects model fitted to cross-classified categorical data, possible explanations for lack-of-fit are an incorrectly specified functional form, omission of important unmeasured covariates (i.e. overdispersion) or absence of important interaction terms. It is generally thought that an additive hazards model is most appropriate for population-based cancer survival analysis although, where lack-of-fit is evident, we may wish to consider a multiplicative model. Overdispersion could occur if, for example, survival depended heavily on some patient characteristic or tumour characteristic (e.g. stage or histology) for which information was not available to us. We have stratified by stage so the most likely explanation for the lack-of-fit is that one or more interaction terms are required. Interactions are required when one or more of the excess hazard ratios vary according to the level of one of the other covariates (i.e. there is effect modification). It is possible, for example, that the excess hazard ratios for age differ according to calendar period of diagnosis, but experience has shown that the most common interaction is between age and follow-up time (for models fitted to data for all stages, the most common interaction is between stage and follow-up time).

An interaction between age and follow-up time means that the excess hazard ratios for age differ according to follow-up time. In other words, an assumption of proportional excess hazards is not appropriate for age. This assumption of proportional excess hazards is rarely justified for stage or age for population-based cancer registry data. Among the group of patients with the worst prognosis, such as those with metastases at diagnosis or those who are elderly at diagnosis, it is common for a large proportion to die very soon after diagnosis (e.g. during the first year), but those who survive the first year have comparably good survival.

Including an age by follow-up interaction term involves the estimation of 12 additional parameters which are, as a group, highly statistically significant using the likelihood ratio test ($G^2 = 59$ for model (1) and $G^2 = 55$ for model (2) on 12 df). Once the interaction term is included there is no longer evidence of lack-of-fit; the deviance is 61 on 58 residual degrees of freedom for model (1) and 58 on 58 residual degrees of freedom for model (2). The parameter estimates associated with sex and calendar period change very little on fitting the interaction term. The estimated excess hazard ratios for age, however, change markedly (Table II). The estimated excess hazard ratio for individuals aged 75+ at diagnosis compared

to patients aged 0−44 is significantly greater than one during the first interval but then less than one during all subsequent intervals. The fitted main effects model erroneously assumes that the excess hazard ratio is identical (e.g. 1.34 for model (4)) for each of the five follow-up intervals (Table I). We have tested the hypothesis of proportional excess hazards against a general alternative hypothesis. A test against a more specific alternative hypothesis, such as a hazard ratio which changes monotonically with time, would be more powerful if the departure from proportional excess hazards was of the form specified by the alternative hypothesis.

Inference based on the model without an interaction term may give the erroneous impression that there is no strong association between age at diagnosis and excess mortality. The truth is that patients aged 75 years or more at diagnosis experience considerably higher excess mortality during the first year following diagnosis; an estimated 3.3 times higher excess mortality than patients aged 0−44. After the first year, however, the excess mortality experienced by the elderly patients is no worse than that experienced by patients of other ages who survive a comparable time. This effect may be a result of the elderly patients being adversely affected by the treatment (or the combined effects of the disease and its treatment) to a greater extent than the younger patients. Among the elderly patients who are sufficiently strong to survive the initial effects of the disease and its treatment, however, the prognosis is comparatively good.

Bolard *et al.* [17] fitted similar models for 2075 colon cancer patients diagnosed in the Côte-d'Or administrative region of France and found evidence of non-proportional excess hazards by stage, age and period of diagnosis. They observed a very similar pattern of excess hazard ratios for age to that shown in Table II.

7. DISCUSSION

Each of the four approaches for estimating the relative survival model described in this paper produce very similar estimates, which is not surprising since they estimate the same underlying model using similar methods. The full-likelihood approach described by Estève *et al.* was hailed as being theoretically superior to the approach described earlier by Hakulinen and Tenkanen since it utilizes information on exact survival times and does not rely on a binomial assumption. Although this is indisputably true, the advantages are minor and the differences in the resulting estimates are small. The choice of approach to use in practice depends not only on theoretical considerations but also on how easy the approach is to apply using available software. A major disadvantage of the Estève *et al.* approach was that it was not previously possible, using the available software package [19], to assess goodness-of-fit, study regression diagnostics, or, most importantly, estimate interaction terms which are required to control for non-proportional excess hazards.

We have showed that by writing the likelihood in terms of subject bands, the model can be easily estimated using a full-likelihood approach in standard software such as SAS, Stata or S-plus and time by covariate interaction terms can be estimated to model non-proportional excess hazards. Regression diagnostics or goodness-of-fit tests are not, however, available since the underlying theory has not been developed. A preferable approach, however, is to estimate the model in the framework of generalized linear models. This retains all of the advantages of the full-likelihood approach (the estimates are, in fact, identical) but brings the additional advantage that regression diagnostics and goodness-of-fit tests are available

(although only when the data are collapsed or grouped). Interaction terms, including time by covariate interaction terms, can be estimated in each of the approaches described in this paper.

Of all the approaches, we recommend the generalized linear model based on collapsed data using exact survival times and a Poisson assumption (approach 4) since one can utilize the theory of generalized linear models for assessing goodness-of-fit and studying regression diagnostics. Implementation of this approach is more user friendly in some software packages than implementation of the full-likelihood approach. For example, categorical predictors can be modelled in SAS PROC GENMOD without requiring the user to create dummy variables explicitly (as must be done before applying the full-likelihood approach). This, of course, is a function of the software rather than the underlying theory. Estimation based on collapsed data is also much faster than estimation, based on individual subject-band data. Estimating the model based on grouped data requires application of the actuarial assumption and approximation of person-time at risk although these approximations are usually reasonable in practice. Clearly, exact survival times should be utilized in preference to grouped survival times whenever possible.

Each approach requires the user to subdivide follow-up into pre-specified intervals and the excess hazard is assumed constant within these intervals. We chose to use annual intervals in our example although intervals of any length may be used with any of the four approaches. The intervals do not have to be of equal length and there is a good argument for using shorter intervals (e.g. of length 3 or 6 months) early in the follow-up where most deaths occur and the excess hazard changes most rapidly. All approaches described in this paper can be used to model relative survival estimated using the so-called period method [26, 27].

As with any model where effects are assumed to be multiplicative (e.g. logistic regression, Poisson regression, Cox regression), one must be aware that relative measures (relative risks) do not always provide the complete picture. It is possible that a large, highly statistically significant relative risk (excess hazard ratio) is of little clinical interest when absolute risks are small. Absolute risks (e.g. excess deaths per 1000 person-years) can easily be estimated based on the fitted model.

We have developed SAS and Stata code (available from http://www.pauldickman.com/rsmodel/) to estimate relative survival using life-table methods and fit the relative survival model using each of the approaches described in this paper. An ongoing topic of discussion among researchers in the field of cancer survival analysis has been the relative merits of the two competing 'models' for relative survival [15, 16] and the merits of the corresponding software packages [19, 28]. Much of the discussion has confused the merits of the 'models' with the merits of the associated software. With this paper we hope that we have resolved much of this controversy by showing that the two underlying models are, in fact, identical and that the different approaches to estimation produce results which are very similar in practice. Utilizing an approach analogous to that used in the analysis of cohort studies, we have shown how the relative survival model can be estimated easily using mainstream statistical software packages, thereby removing the reliance on special-purpose software.

REFERENCES

1. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer* 1977; **35**:1–39.

2. Bull K, Spiegelhalter DJ. Tutorials in biostatistics: survival analysis in observational studies. *Statistics in Medicine* 1997; **16**:1041–1074.
3. Gail MH. Competing risks. In: Kotz S, Johnson NL (eds), *Encyclopedia of Statistical Sciences*, vol. 2. Wiley: New York, 1982; 75–81.
4. Tsiatis AA. Competing risks. In: Armitage P, Colton T (eds), *Encyclopedia of Biostatistics*, vol. 1. Wiley: Chichester, 1998; 824–834.
5. Percy CL, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American Journal of Public Health* 1981; **71**:242–250.
6. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *National Cancer Institute Monograph* 1961; **6**:101–121.
7. Berkson J, Gage RP. Calculation of survival rates for cancer. *Proceedings of Staff Meetings of the Mayo Clinic* 1950; **25**:270–286.
8. Estève J, Benhamou E, Raymond L. *Statistical Methods in Cancer Research Volume IV*: *Descriptive Epidemiology.* IARC Scientific Publications No. 128. International Agency for Research on Cancer: Lyon, 1994.
9. Oksanen H. Modelling the survival of prostate cancer patients. *Ph.D. Thesis*, University of Tampere, Finland, 1998 (Acta Universitatis Tamperensis, vol. 615).
10. Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Edwards BK (eds). *SEER Cancer Statistics Review*, *1973–1999*. National Cancer Institute: Bethesda, 2002. http://seer.cancer.gov/csr/1973_1999/.
11. Coleman MP, Babb P, Damiecki O. Grosclaude P, Honjo S, Jones J, Knerer G, Pitard A, Quinn M, Sloggett A, De Stavola B. *Cancer Survival Trends in England and Wales*, *1971–1995*: *Deprivation and NHS Region.* No. 61 in Studies in Medical and Population Subjects. The Stationary Office: London, 1999.
12. Berrino F, Capocaccia R, Estève J, Gatta G, Hakulinen T, Micheli A, Sant M, Verdecchia A. (eds). *Survival of Cancer Patients in Europe*: *The EUROCARE-2 Study*. IARC Scientific Publications No. 151. International Agency for Research on Cancer: Lyon, 1999.
13. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B* 1972; **34**:187–220.
14. Buckley JD. Additive and multiplicative models for relative survival rates. *Biometrics* 1984; **40**:51–62.
15. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Applied Statistics* 1987; **36**:309–317.
16. Estève 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.
17. Bolard P, Quantin C, Estève J, Faivre J, Abrahamowicz M. Modelling time-dependent hazard ratios in relative survival: application to colon cancer. *Journal of Clinical Epidemiology* 2001; **54**:986–996.
18. Suissa S. Relative excess risk: an alternative measure of comparative risk. *American Journal of Epidemiology* 1999; **150**:279–282.
19. Hédelin G. *RELSURV 2.0 a Program for Relative Survival Analysis*. Department of Epidemiology and Public Health, Faculty of Medicine, Louis Pasteur University, Strasbourg, France, 1997.
20. Andersen PK, Borgan, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: Berlin, 1995.
21. Breslow NE, Day NE. *Statistical Method in Cancer Research*: *Volume II – The Design and Analysis of Cohort Studies*. IARC Scientific Publications No. 82. International Agency for Research on Cancer: Lyon, 1987.
22. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall: London, 1989.
23. MacLennan R. Items of patient Information which may be collected by registries. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG (eds), *Cancer Registration*: *Principles and Methods*, IARC Scientific Publications No. 95. International Agency for Research on Cancer: Lyon, 1991; 43–63.
24. Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford University Press: Oxford, 1993.
25. Dickman PW, Hakulinen T, Luostarinen T, Pukkala E, Sankila R, Söderman B, Teppo L. Survival of cancer patients in Finland 1955–1994. *Acta Oncologica* 1999; **38**(Suppl. 12):1–103.
26. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996; **78**: 2004–2010.
27. Brenner H, Hakulinen T. Up-to-date long-term survival curves of patients with cancer by period analysis. *Journal of Clinical Oncology* 2002; **20**:826–832.
28. Voutilainen ET, Dickman PW, Hakulinen T. *SURV2*: *Relative Survival Analysis Program.* Finnish Cancer Registry: Helsinki, 2000. http://www.cancerregistry.fi/surv2/.