STATISTICS IN MEDICINE, VOL. 9, 529–538 (1990)

RELATIVE SURVIVAL AND THE ESTIMATION OF NET SURVIVAL: ELEMENTS FOR FURTHER DISCUSSION

J. ESTÈVE

International Agency for Research on Cancer, 150, cours Albert-Thomas, 69372 Lyon Cedex 08, France

E. BENHAMOU

Institut Gustave-Roussy, 39, rue Camille Desmoulins, 94805 Villejuif Cedex, France

M. CROASDALE

Central Electricity Generating Board, Health and Safety Department, Courtenay House, 18 Warwick Lane, London EC4P 4EB, U.K.

AND

L. RAYMOND

Registre genevois des Tumeurs, 55, bd de la Cluse, 1205 Geneva, and Department of Preventive and Social Medicine, University of Geneva, Switzerland

SUMMARY

The methods of calculation of survival corrected for independent cause of death are discussed, and a maximum likelihood method is proposed and illustrated by survival of colon cancer patients in Geneva. The methods which are at present favoured for doing such calculations are subject to various biases when estimating net survival if the populations are heterogeneous for life expectancy. The proposed maximum likelihood approach would eliminate these biases by enabling relevant adjustment for covariates which influence survival. The routine use of such methods would permit better comparison of survival within and between populations.

1. INTRODUCTION

The statistical literature on survival analysis is most often associated with the randomized clinical trial. Indeed many important methodological results related to the analysis of censored observations have originated from continuous collaboration between clinicians and biostatisticians working on the evaluation of cancer therapy. Comparatively less attention has been given to the study of cancer patient survival itself, though the estimation and comparison of population-wide survival is of the greatest importance for public health authorities. In particular, it may help to evaluate the effectiveness of the health care system in various segments of the population, and may eventually permit study of the net effect of therapeutic improvements on the population as a whole. Although the basic survival methodology is similar to that used in clinical trials, the problems of interpretation which arise in such group description and comparison are essentially different.[1] One of the tools for undertaking such a task is relative survival,[2] which has been used widely though unfortunately not always understood. This important concept was created to

0277–6715/90/050529–10$05.00

Received April 1989
Revised October 1989

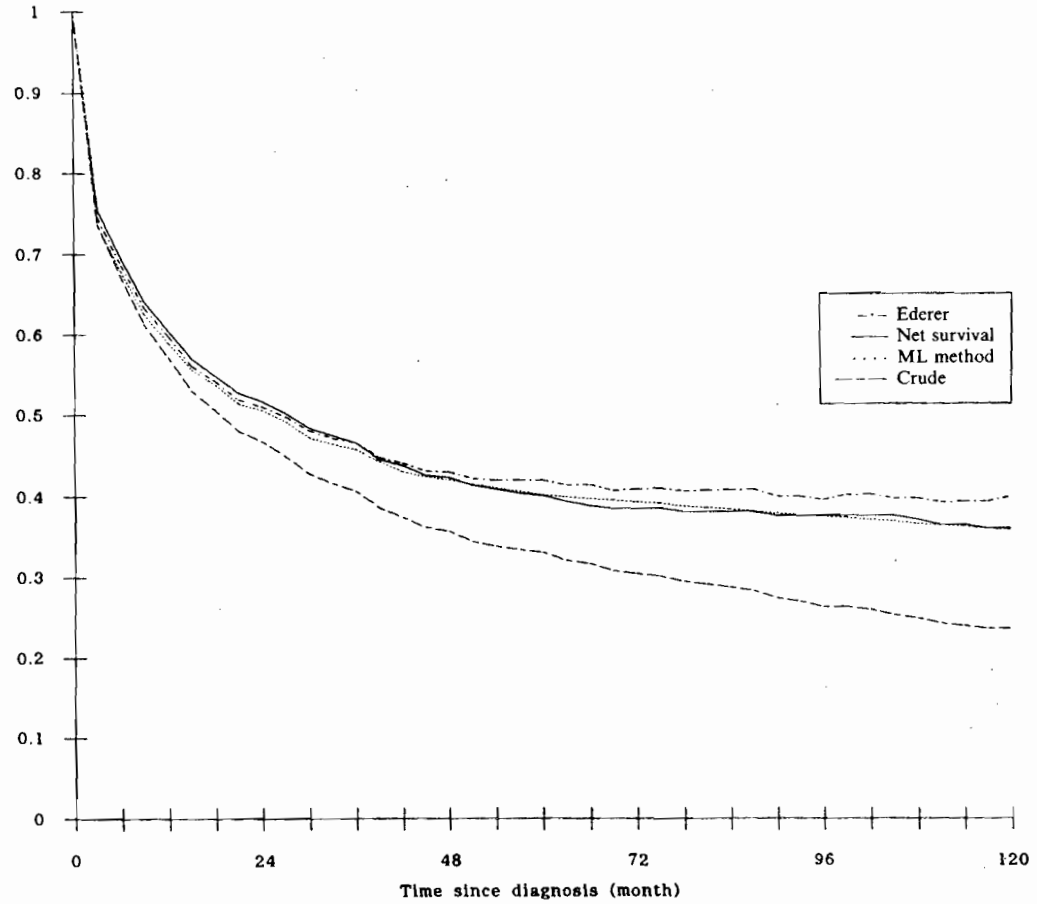

Figure 1. Survival of colon cancer patients, both sexes, Geneva 1970–79

provide an objective measure of the proportion of patients dying from the direct or indirect consequences of disease in a given population, and hence a measure of patient survival corrected for the effect of other independent causes of death.[3] In evaluating the survival of a group of cancer patients, many clinicians would exclude from their calculations deaths unrelated to cancer, considering it unfair to attribute to the disease those deaths which would have occurred in its absence. A practical evaluation would consider non-cancer deaths as censored observations in the usual actuarial or product limit estimates of survival; the resulting curve would represent the net effect of cancer on survival, and is known as the *net survival curve* (Figure 1). This curve tends to level off after several years and the corresponding limit is usually interpreted as the probability of being cured of the disease. However, it has long been recognized[3] that there is no fully satisfactory method of deciding whether a given death should be classified as a cancer or as a 'non-cancer' death. Even if such a method existed, the available information on cause of death (recorded on the death certificate or in other routinely collected information) would rarely be sufficiently accurate to make such a classification possible when dealing with population data. Given this lack of

information, it was proposed[2, 3] to rely on the vital statistics available for the population under study to estimate net survival. The purpose of this article is to discuss the methodology related to this later approach which resulted in the notion of *relative survival.*

2. RELATIVE SURVIVAL

Relative survival is the ratio of the observed survival of a given cohort of (cancer) patients to the survival that the group should have experienced based on the life table of the population from which they were diagnosed.[2] Usually, the cohort of patients will include all cancers of a given site identified at a cancer registry, and expected survival will be calculated from the life table of the population covered by the registry. Although this is hardly explicit in the above reference, the intention of the originators of this concept was to estimate net survival since

$$S_o(t) = S_c(t)S_e(t) \tag{1}$$

is equivalent to

$$\lambda_o(t) = \lambda_c(t) + \lambda_e(t), \tag{2}$$

where S and λ stand for survival and hazard respectively, and o, c, and e for observed (in fact observable), corrected and expected. The method of estimation of $S_c(t)$ has been questioned on several grounds, and Ederer himself proposed two versions of his method based on life tables. In the first[2] the probabilities of surviving t years after diagnosis are obtained from the relevant life tables for all individuals in the study cohort and summed to get the expected number of survivors after t years; the expected survival is obtained by dividing by the initial size of the cohort. Relative survival is then estimated as the ratio of the observed survival (actuarial) and the expected survival. In the second method,[4] the probability of dying in the next 'actuarial' interval is estimated at the beginning of the interval for each individual still at risk at that time. The sum of these probabilities gives the expected number of deaths in the interval. An expected actuarial survival may then be calculated, and the relative survival at t years after diagnosis is the ratio of the observed actuarial survival and the expected actuarial survival at this time point. The two methods are evidently different since in the second expected survival at t years is dependent on the observed survival and on the pattern of withdrawal.

The main difficulty of this methodology was identified a long time ago by Hakulinen,[5] and relates to heterogeneity within the study population of covariates which affect survival, and to changes in their distribution as the study group is depleted by death from the disease and from other causes. Age at diagnosis is an important example of such a covariate; if we suppose that net survival depends on age x, then relative survival as defined in the first approach[2] would estimate

$$\bar{S}_c(t) = \frac{\sum_{x=1}^{g} n_x S_e(t, x) S_c(t, x)}{\sum_{x=1}^{g} n_x S_e(t, x)}, \tag{3}$$

where the numerator is the 'observable' survival of the study cohort and $S_c(t, x)$ denotes net survival in the age group x comprising initially n_x individuals. $\bar{S}_c(t)$ is therefore a weighted average of net survival, the weights being proportional to the number of survivors which would be expected in the various subgroups at time t, based on the life table of the general population. As noted by Hakulinen,[5] this statistic will estimate in the long term the net survival of the group with the longest life expectancy. Since for some cancers net survival is better for those diagnosed at a younger age, it is not surprising to observe that the 'empirical' relative survival increases with

time, simply as a result of the increasing dominance of these individuals and not, as has been suggested, because cancer survivors receive better medical care than the general population; it is therefore important to recognize that relative survival is heavily influenced by the initial age composition of the cohort and by its evolution with time. Consequently, it is difficult to use this statistic for comparative purposes between populations, and several alternatives have been proposed (see Hakulinen *et al.*[6] for a recent review).

In the next section we describe a full maximum likelihood solution, and we place this approach in the context of other proposed alternatives in the discussion.

3. A MAXIMUM LIKELIHOOD SOLUTION

This method has been suggested by several authors[7–9] but has never been entirely worked out; we propose here a means of implementation.

The mortality rate of the cohort under study at time t after diagnosis may be written

$$\lambda_o(t, z, x) = \lambda_c(t, z) + \lambda_e(x + t, z_1), \tag{4}$$

where x is age at diagnosis; z is a vector of covariates; z_1 is a subvector of z for which mortality from other causes is available (for example sex, ethnic group); λ_e is the net hazard to be estimated; and λ_e is a known hazard function, usually obtained from published vital statistics.

Estimation of λ_e by maximum likelihood is greatly facilitated if we represent this function by a parametric model. The choice of model here follows the present trend of using the Cox regression model for calculating survival of large cohorts. Given the number of individuals in the cohort and the inevitable lack of precision in the time variable, many ties will occur and the approximation to the likelihood which is used is equivalent to considering the hazard function constant within post-specified time intervals.[10] Therefore we suggest that the intervals be pre-specified and survival be calculated using either the actuarial method, the Cox regression model or maximum likelihood for net survival, according to the purpose of the calculation. This policy corresponds to present usage in cancer registries, where there is some preference for grouped data, and places all these calculations within a common framework. We shall then assume a multiplicative model of the form

$$\lambda_c(t, z) = \exp(\beta z) \sum_{k=1}^{m} \tau_k I_k(t), \tag{5}$$

where $I_k(t)$ is the indicator function for the kth interval and τ_k is the net mortality rate in that interval for patients with $z = 0$. The statistical problem consists of estimating β and τ, using the maximum likelihood method, from a sample of n observations t_i, δ_i, z_i and x_i, $i = 1, \ldots, n$, where t_i is the survival time and δ_i denotes as usual the censoring index (1 if t_i is the time of death; 0 if t_i is a censored observation). The log-likelihood may be written (up to a constant)

$$L(\beta, \tau) = -\sum_{i=1}^{n} \Lambda_c(t_i, z_i) + \sum_{i=1}^{n} \delta_i \log[\lambda_c(t_i, z_i) + \lambda_e(t_i + x_i, z_{1i})], \tag{6}$$

and the first and second derivatives are given by

$$\frac{\partial L}{\partial \beta_j} = \sum_{k=1}^{m} \tau_k \sum_{i=1}^{n} \frac{z_{ji} R_i}{\lambda_c + \lambda_e} [\delta_i I_k(t_i) - (\lambda_c + \lambda_e) t_{ki}] \tag{7}$$

$$\frac{\partial L}{\partial \tau_k} = \sum_{i=1}^{n} \frac{R_i}{\lambda_c + \lambda_e} [\delta_i I_k(t_i) - (\lambda_c + \lambda_e) t_{ki}] \tag{8}$$

$$\frac{\partial^2 L}{\partial \beta_j \partial \beta_l} = \sum_{i=1}^{n} z_{ji} z_{li} \left[\frac{\delta_i \lambda_c \lambda_e}{(\lambda_c + \lambda_e)^2} - \sum_{k=1}^{m} R_i \tau_k t_{ki} \right] \tag{9}$$

$$\frac{\partial^2 L}{\partial \tau_k \partial \beta_j} = \sum_{i=1}^{n} z_{ji} R_i \left[\frac{\delta_i I_k(t_i) \lambda_e}{(\lambda_c + \lambda_e)^2} - t_{ki} \right] \tag{10}$$

$$\frac{\partial^2 L}{\partial \tau_k \partial \tau_h} = - \sum_{i=1}^{n} \delta_i I_k(t_i) I_h(t_i) \frac{R_i^2}{(\lambda_c + \lambda_e)^2}, \tag{11}$$

where $\Lambda_c(t_i, z_i)$ is the net cumulative hazard rate up to t_i; λ_c and λ_e stand respectively for $\lambda_c(t_i, z_i)$ and $\lambda_e(t_i + x_i, z_{1i})$; z_{ji} is the value of the jth covariate for the ith individual; $R_i = \exp(\beta z_i)$; and $t_{ki} = \int_0^{t_i} I_k(t)\,dt$ is the time spent by the ith patient in the kth interval.

The ML procedure appears therefore (see equations (7) and (8)) as a refinement of a method which consists of subtracting the number of expected deaths from the observed deaths in each interval and inferring the mortality rate from this adjusted number. We may then foresee some difficulties when the number of expected deaths is already greater than the observed deaths. This will occur when very few people remain alive or when the pre-specified intervals are too small. The practical solution is to constrain the τ_k to be positive and to choose intervals in such a way that it will be unlikely that the maximum of the likelihood will be reached on the boundary of the valid domain ($\tau_k = 0$). However, it should be kept in mind that τ_k may truly tend to zero as k increases since, for some sites at least, a proportion of individuals are cured of disease.

To help the user of the method, a program (available on request to the first author) makes a rough estimate of relative survival for two dates and fits a log-logistic model to evaluate the probability that a given interval will provide sufficient deaths to estimate the net hazard rate. Starting with 3 month intervals, the program groups them together in such a way that (i) the number of observed deaths is likely to exceed the number of expected deaths in each interval, and (ii) the length of successive intervals increases and corresponds to a partition of each year of follow-up if shorter than 1 year. Users may alternatively build their own partition. This lengthy strategy is employed to avoid a data driven choice which would be criticized on the basis of well-known statistical arguments. In the example presented below the follow-up was 10 years, and this strategy led to a partition in sixteen intervals: twelve intervals of 3 months up to 3 years; two intervals of 6 months up to 4 years; one interval of 1 year; and one interval of 5 years. This partition was retained in the analysis. As noted above, the net hazard should tend to zero, and this strategy is bound to define longer intervals at the end of the follow-up, owing to the decreasing number of deaths caused by the cancer under study. To test whether the net survival rate tends to zero, one could look at the confidence interval of the last τ, or alternatively dichotomize the last interval and test the hypothesis that the last τ so created is equal to zero using the likelihood ratio test.

4. EXAMPLE

The Cancer Registry at Geneva, which was created in 1970, made available to the authors anonymous data on 940 colon cancer patients (ICD = 153) registered between 1970 and 1979. The vital status of each registered cancer case is ascertained routinely on the 5th, 10th and 15th anniversaries after diagnosis and, when relevant, the cause of death is validated by a physician working at the registry;[11] only for 2 per cent of these patients was follow-up incomplete. Therefore it was possible to calculate relative and net survival using the life table for Switzerland (1978–83); net survival was also obtained from the Cox method considering non-cancer deaths as censored observations. The cohort was followed up to 1984 and consists of 454 men and 486

Table I. Survival of colon cancer patients, Geneva 1970–79: age and sex effects on crude and net survival

Proportional hazard model	Parameter estimates§ (standard error) Sex (men)	Age (years) 65–74	75 +	Log likelihood
Crude survival*	0·101	0·621	0·986	− 4006·28
	(0·08)	(0·11)	(0·10)	
Test for sex	—	0·616	0·970	− 4007·07
		(0·11)	(0·10)	
Net survival†				
(life table estimate)	0·048	0·593	0·916	− 1039·89
	(0·09)	(0·12)	(0·12)	
Test for sex	—	0·591	0·909	− 1040·02
		(0·12)	(0·12)	
Net survival‡				
(cause-specific estimate)	0·005	0·613	0·934	− 3364·78
	(0·09)	(0·12)	(0·11)	
Test for sex	—	0·612	0·934	− 3364·78
		(0·12)	(0·11)	

* Survival from all causes. The censored observations are those of people withdrawn and lost to follow-up. Cox regression model with a time unit of 3 months is used for estimation.
† Survival corrected for other causes as described by equations (4) and (5). Maximum likelihood estimates as described in this article.
‡ Net survival estimated by considering 'non-cancer' deaths as censored observations. Cox regression model with a time unit of 3 months is used for estimation.
§ The exponential of these figures gives the estimate of the relative hazard rate for the specified groups.

Table II. Net survival of colon cancer patients, Geneva 1970–79: comparison of various methods of estimation

Age (years)	No. cases	Life table method ML method* 5 years	ML method* 10 years	Ederer method† 5 years	Ederer method† 10 years	Cause-specific net survival‡ 5 years	Cause-specific net survival‡ 10 years
< 65	322	0·56	0·51	0·54	0·50	0·57	0·55
		(0·03)	(0·04)	(0·03)	(0·04)	(0·03)	(0·03)
65–74	292	0·35	0·30	0·37	0·30	0·38	0·31
		(0·03)	(0·04)	(0·03)	(0·05)	(0·03)	(0·04)
75 +	326	0·24	0·19	0·32	0·30	0·25	0·21
		(0·03)	(0·03)	(0·03)	(0·09)	(0·03)	(0·03)
Total	940	0·40	0·36	0·42	0·40	0·40	0·36
		(0·02)	(0·03)	(0·02)	(0·03)	(0·02)	(0·02)

* As derived from the second model in Table I with the sex effect omitted.
† Calculated by the method described in Reference 2; actuarial method is used with 3-month intervals and the life table of Switzerland (1978–83) is used for the calculation of expected survival.
‡ The cause of death coded at the registry determined whether an observed death is a true response or a censored observation; the actuarial method of calculation was used with 3-month intervals.

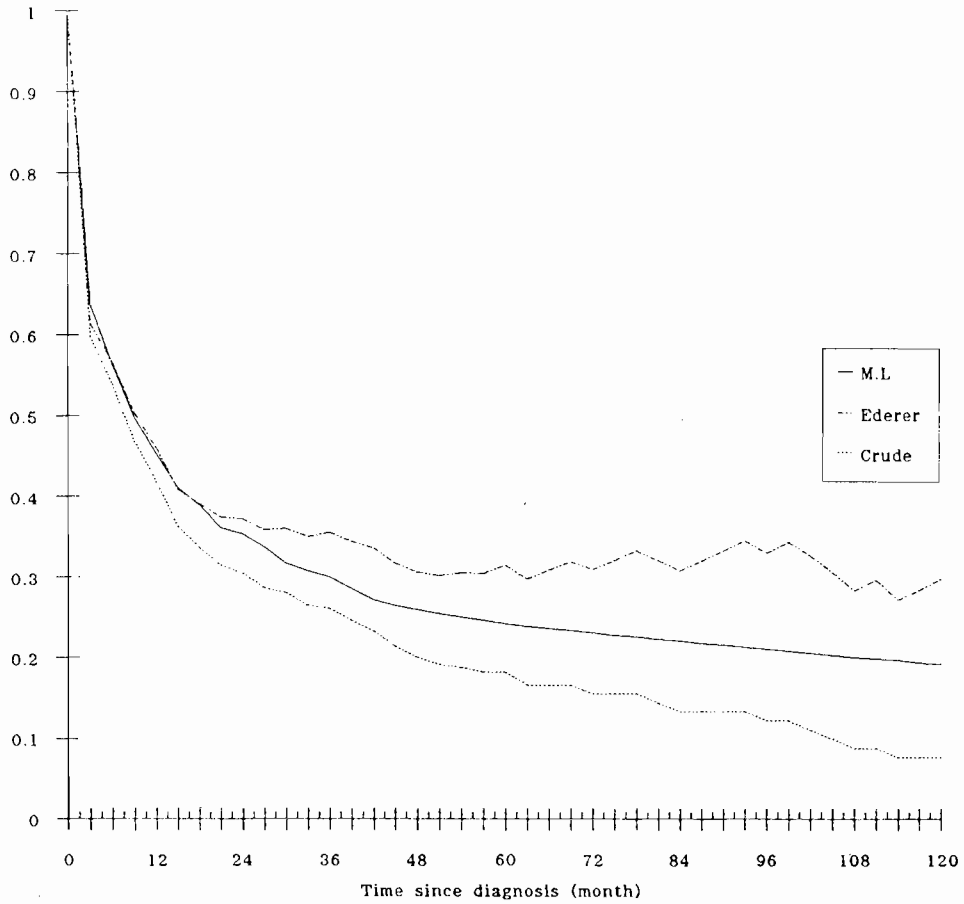

Figure 2. Survival of colon cancer patients, both sexes, age 75 + years, Geneva 1970–79

women. The men were younger (mean age 66) than the women (mean age 68·5). Table I describes the effects of age and sex on survival. The crude survival is slightly worse for men but not significantly so, and net survival is the same in both sexes. There is a very strong age effect on net survival, almost as large as the age effect on crude survival. This effect is slightly stronger when the computation is done from the cause of death, suggesting that there is a tendency to attribute death to other causes in younger people. This is also suggested in Table II, where it may be seen that net survival calculated from the life table is marginally lower than net survival obtained from the cause of death. This discrepancy could also be explained by some inadequacy of the Swiss life table for the city of Geneva. It is also evident from Table II and Figure 2 that the Ederer method is in this particular instance too optimistic for the oldest age group. Furthermore, this method gives the wrong impression that survival is constant after 5 years (Figures 1 and 2), which is contradicted by the mean hazard rate τ in the interval 5 to 10 years being equal to 0·019 per year with a standard error of 0·008. Division of the last time interval into two intervals of 30 months results in estimates of τ (and its standard error) of 0·022 (0·010) and 0·013 (0·012) respectively, with a likelihood of $-$ 1039·9, which is almost the same as previously. Although the net survival rate is no longer significantly positive after 7·5 years, there is no indication that the exponential survival

model for the last 5 year interval is oversimplistic. Finally, we should note that in this particular example the life table and the cause-specific methods give exactly the same value for the net survival of the whole cohort (Table II and Figure 1).

5. DISCUSSION

In this article we have proposed a maximum likelihood method for computing net survival when the causes of death of cohort members are not known or when it seems advisable not to use them. The method is well adapted to population data and has a natural place in the statistical methodology for survival computation. The favourite method for calculating relative survival[2] has several shortcomings when estimating net survival, especially for populations which lack homogeneity in covariates which influence either net survival or survival from other causes. Hakulinen[5, 12] discussed this problem and proposed a correction for heterogeneity in patient withdrawal, but retained the basic idea of calculating an expected survival which depends only on the initial composition of the cohort and on the potential time of follow-up. He argued that this expected survival should not be influenced by the way the cohort is modified by the cause of death under study as in the method proposed by Ederer and Heise[4] and programmed by Rothman and Boice.[13] However, none of these methods avoids the propensity to overestimate long-term net survival in a group which has heterogeneous life expectancies. In the oldest age group of the Geneva cohort, for example, the Ederer and Heise estimate of 10 year survival was 0·24 and the estimate obtained with the method of Hakulinen was 0·30. As suggested by the last author, one should distinguish clearly between relative survival (the ratio of observed to expected survival) and net survival as estimated from modelling the hazard rate. These two concepts could coincide in some situations; but in other contexts, observed differences between them provide some information about the pattern of survival in the study group.

Breslow[14] suggested calculating an SMR for the cohort of patients as if the presence of disease acted multiplicatively on the hazard rate of the general population. This idea was later developed within a more general discussion of multiplicative and additive modelling of excess mortality, and it was emphasized that the SMR is the appropriate choice only in the framework of the multiplicative model.[15, 16] Hill *et al.*[17] calculated the expected deaths in each follow-up interval from a multiplicative model and constructed an expected survival curve from these expected deaths. Their method is therefore similar to that of Ederer and Heise.[4] Pocock *et al.*[8] used basically the same methodology as Ederer and Heise to calculate long-term relative survival for pre-menopausal and post-menopausal breast cancer women in Scotland. However, they supported an additive hazard model similar to that used in the present paper, but supposed that the excess death rate decreased exponentially with time. Andersen *et al.*[18] improved considerably the use of the multiplicative model in this context. They rightly employed age as the time metameter and observed that the problem can be reduced to the estimation of a Cox proportional hazard model with time dependent covariates. The application they had in mind was the modelling of the survival of insulin-dependent diabetics, and this may explain their choice. The excess death rate due to this disease should have a very different shape from that caused by cancer and could well be described correctly in the framework of a multiplicative model. The problem of modelling and estimating excess and relative mortality is further discussed in a research report of Andersen and Vaeth,[19] following ideas of Aalen;[20] in this report they proposed to estimate the cumulative excess mortality from the difference between the observed cumulative hazard and an expected cumulative hazard obtained from the known background risk in the general population. This approach would be again a direct extension to continuous data of the method of Ederer and Heise.[4] It is clear from this discussion that, although the method of Ederer *et al.*[2] is at present the

method of choice for routinely published survival data from cancer registries, there is little agreement on its superiority for estimating net survival.

The computation of net survival often has comparative purposes, and it is appropriate to use it if the populations of patients to be compared have different 'normal' life expectancy. If this is not so, then a simple comparison of raw survival adjusted for age may be sufficient. The present maximum likelihood method may also be applied here, and some advantage could come from prior knowledge that the 'competing cause' age effect acts additively;[15] it may therefore be useful to separate it from the multiplicative effect acting on net survival. When net survival depends on age, the two approaches above are superior to the simple comparison of relative survival since, as noted by Ederer *et al.*[2] but often forgotten subsequently, relative survival estimates are not automatically adjusted for age (see also Myers and Hankey[21] and Hakulinen[1]). When different life tables should be used for the populations compared, the methodology described in the present article is appropriate and easy to apply. It has in fact been demonstrated that this type of piecewise exponential model has useful theoretical properties which could probably be extended in the present context.[22] Even though proportional hazard models have been criticized on the grounds of complexity or because the hypothesis of proportionality is not realistic,[7] we should nevertheless remember that they provide at least the same power of analysis as the more simplistic methods. An analysis with different hazard rates in strata is always possible and, except for the choice of interval, the same strategy as for the Cox regression analysis would apply to this ML method for net survival. Other methods have been proposed, and are based on the comparison of the number of deaths between groups in each follow-up interval. Brown[23] developed an extension of the Mantel–Haenszel approach in which relative survival appears as a nuisance parameter. The corresponding score test is a weighted sum of differences between observed and expected deaths based on the estimated common relative survival. Hakulinen *et al.*[6, 24] modelled the expected deaths in each follow-up interval using a generalized linear model. It is to be noted that these two methods hypothesized a binomial distribution for the number of observed deaths in each follow-up interval. This is not usually the case; on the contrary, at the beginning of each interval the population is heterogeneous for expected survival and the probability of surviving to the end of it is different from person to person. This is taken into account in Hakulinen's method in allowing different dispersion from that anticipated by the binomial distribution; he suggested estimating the true variance with a scale parameter in the GLIM macro proposed for undertaking such an analysis.[25]

The analysis of survival data from cancer registries will be more frequent in the future, and it is important that a robust and standard methodology is used everywhere. We have proposed here an approach which, in our opinion, has some advantages over previous methods. Other methodological problems, related to the availability of the stage of diagnosis or the procedure of follow-up, should be considered[26] before routinely obtained figures can be published as is presently done for incidence data.[27]

ACKNOWLEDGEMENTS

This research was carried out while Martin Croasdale was on sabbatical leave at the IARC. We are grateful to the CEGB for granting him permission to participate in this work. The authors also thank the group of Latin cancer registries which brought this problem to their attention on the occasion of one of its methodological seminars, and Drs. Hakulinen and Kaldor for useful comments on a first version of the paper.

REFERENCES

1. Hakulinen, T. 'A comparison of nationwide cancer survival statistics in Finland and Norway', *World Health Statistics Quarterly*, **36**, 35–46 (1983).

2. Ederer, F., Axtell, L. M. and Cutler, S. J. 'The relative survival rate: a statistical methodology', *National Cancer Institute monographs*, **6**, 101–121 (1961).
3. Berkson, J. and Gage, R. P. 'Calculation of survival rates for cancer', *Proceedings of the Staff Meetings of the Mayo Clinic*, **25**, 270–286 (1950).
4. Ederer, F. and Heise, H. 'The effect of eliminating deaths from cancer on general population survival rates', methodological note 11, End Results Evaluation Section, National Cancer Institute, August 1959.
5. Hakulinen, T. 'On long-term relative survival rates', *Journal of Chronic Diseases*, **30**, 431–443 (1977).
6. Hakulinen, T., Tenkanen, L. and Abeywickrama, K. 'Testing equality of relative survival patterns based on aggregated data', *Biometrics*, **43**, 313–325 (1987).
7. Buckley, J. D. 'Additive and multiplicative models for relative survival rates', *Biometrics*, **40**, 51–62 (1984).
8. Pocock, S. J., Gore, S. M. and Kerr, G. R. 'Long term survival analysis: the curability of breast cancer', *Statistics in Medicine*, **1**, 93–104 (1982).
9. Andersen, P. K., Borch-Johnsen, K., Decker, T., Green, A., Hougaard, P., 'Keiding, N. and Kreiner, S. 'A Cox regression model for the relative mortality and its application to diabetes mellitus survival data', *Biometrics*, **41**, 921–932 (1985).
10. Breslow, N. E. 'Contribution to the discussion on the paper by D. R. Cox, "Regression models and life tables"', *Journal of the Royal Statistical Society Series B*, **34**, 216–217 (1972).
11. Raymond, L., Obradovic, M. and Fioretta G. 'Facteurs pronostiques de survie après cancer: le modèle épidémiologique', *Médecine Sociale et Préventive*, **33**, 269–273 (1988).
12. Hakulinen, T. 'Cancer survival corrected for heterogeneity in patient withdrawal', *Biometrics*, **38**, 933–942 (1982).
13. Rothman, K. J. and Boice, J. D. 'Epidemiologic analysis with a programmable calculator', NIH Publication 79-1649, US Department of Health, Education and Welfare, Bethesda, 1979.
14. Breslow, N. E. 'Analysis of survival data under the proportional hazards model', *International Statistical Review*, **43**, 45–58 (1975).
15. Breslow, N. E. and Day, N. E. 'The standardized mortality ratio', in Sen, P. K. (ed.), *Statistics in Biomedical, Public Health and Environmental Sciences*, North-Holland/Elsevier Science Publishers, Amsterdam, 1985, pp. 55–74.
16. Breslow, N. E. and Day, N. E. *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies*, IARC Scientific Publication 82, International Agency for Research on Cancer, Lyon, 1987.
17. Hill, C., Laplanche, A. and Rezvani, A. 'Comparison of the mortality of a cohort with the mortality of a reference population in a prognostic study', *Statistics in Medicine*, **4**, 295–302 (1985).
18. Andersen, P. K., Borch-Johnsen, K., Deckert, T., Green, A., Hougaard, P., Keiding, N. and Kreiner, S. 'A Cox regression model for the relative mortality and its application to diabetes mellitus survival data', *Biometrics*, **41**, 921–932 (1985).
19. Andersen, P. K. and Vaeth, M. *Excess and relative mortality*, Research report 87/7, Statistical Research Unit, University of Copenhagen, 1987.
20. Aalen, O. O. 'Non-parametric inference for a family of counting processes', *Annals of Statistics*, **6**, 701–726 (1978)..
21. Myers, M. H. and Hankey, B. F. 'Cancer patient survival experience'. NIH Publication 80-2148, US Department of Health and Human Services, 1980.
22. Frièdman, M. 'Piecewise exponential models for survival data with covariates', *Annals of Statistics*, **10**, 101–113 (1982).
23. Brown, C. C. 'The statistical comparison of relative survival rates', *Biometrics*, **39**, 941–948 (1983).
24. Hakulinen, T. and Tenkanen, L. 'Regression analysis of relative survival rates', *Applied Statistics*, **36**, 309–317 (1987).
25. Hakulinen, T., Gibberd, R., Abeywickrama, K. and Söderman, B. 'A computer program package for cancer survival studies, version 1.0', Cancer Society of Finland Publication 39, Finnish Cancer Registry and University of Newcastle, Australia, 1988, p. 20.
26. Tallis, G. M., Leppard, P. and O'Neill, T. J. 'The analysis of survival data from a central cancer registry with passive follow-up', *Statistics in Medicine*, **7**, 483–490 (1988).
27. Muir, C. S., Waterhouse, J., Mack, T., Powell, J. and Whelan, S. *Cancer Incidence in Five Continents*, Volume V, IARC Scientific Publication 88, 1987.