ELSEVIER

Journal of Health Economics 18 (1999) 365–380

JOURNAL OF HEALTH ECONOMICS

On the use of survival analysis techniques to estimate medical care costs

Ruth D. Etzioni [a,*], Eric J. Feuer [b], Sean D. Sullivan [c], Danyu Lin [d], Chengcheng Hu [d], Scott D. Ramsey [e]

[a] *Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, MP-665, P.O. Box 19024, Seattle, WA 98109-1024, USA*
[b] *National Cancer Institute, USA*
[c] *School of Pharmacy, University of Washington, Seattle WA, USA*
[d] *Department of Biostatistics, University of Washington, Seattle WA, USA*
[e] *Departments of Medicine and Health Services, University of Washington, Seattle WA, USA*

Received 1 August 1997; revised 1 September 1998; accepted 1 September 1998

Abstract

Measurement of treatment costs is important in the evaluation of medical interventions. Accurate cost estimation is problematic, when cost records are incomplete. Methods from the survival analysis literature have been proposed for estimating costs using available data. In this article, we clarify assumptions necessary for validity of these techniques. We demonstrate how assumptions needed for valid survival analysis may be violated when these methods are applied to cost estimation. Our observations are confirmed through simulations and empirical data analysis. We conclude that survival analysis approaches are not generally appropriate for the analysis of medical costs and review several valid alternatives.

JEL classification: C000 Mathematical and quantitative methods: general; C100 Econometric and statistical methods: general; C400 Econometric and statistical methods: special topics: general

Keywords: Survival analysis techniques; Treatment costs; Medical interventions

* Corresponding author. Tel.: +1-206-667-4145; fax: +1-206-667-7004; e-mail: etzi@fhcrc.org

0167-6296/99/$ - see front matter
PII: S0167-6296(98)00056-3

1. Introduction

With the rapid escalation of the costs of medical treatment, interest in accurately quantifying the cost of medical care has increased. Estimates from cost studies are needed to determine the economic burden of disease, to predict the economic consequences of new medical interventions, and for comparative purposes such as cost-effectiveness analysis.

Many studies of medical costs focus on estimating the costs of care over a specific time period such as an episode of illness or a lifetime after diagnosis of disease. In these cases, costs may be thought of as accumulating over time prior to a terminating event. If patients are not followed long enough to experience the event while under observation, they are said to be censored. Economists have recently observed that censoring could lead to biased estimates of costs unless appropriately accounted for in the analysis (Dudley et al., 1983; Fenn et al., 1995, 1996).

Survival analysis is a body of techniques for analyzing lifetimes under censoring. Survival analysis has a long history in economic as well as biomedical application. In the area of labor economics, for example, employment durations are treated as survival times and analyzed accordingly (Heckman and Singer, 1985; Kiefer, 1988; Lancaster, 1990). Recently, survival analysis approaches have been proposed for analyzing medical costs. In the survival analysis approach to cost data, individuals' cumulative costs are treated like survival times and analyzed accordingly (Dudley et al., 1983; Fenn et al., 1995, 1996).

The survival analysis approach to costs seems appealing because of its simplicity, its nonparametric nature, and its apparent robustness in the presence of censoring. However, although the approach is apparently free from distributional assumptions, it is not entirely assumption-free. In this article, we explain the assumptions necessary for a survival analysis to be valid and show how they might be violated when survival analysis is applied directly to possibly censored data on cumulative costs. Using simulation modeling, we demonstrate the impact of the violation of assumptions, first in the standard time domain and then in the cost domain. Finally, we review some alternative, nonparametric methods that have been developed, and show how the results of these methods differ from the results of survival analysis in a real costs dataset.

2. Survival analysis

In this section we review the most commonly used survival analysis techniques for estimating distributions of lifetimes and the association between lifetimes and explanatory covariates. We illustrate concepts first in the time domain and then in the costs domain.

2.1. Censoring and failure

In classical survival analysis, interest focuses on the time to an event, most commonly a failure of some sort. Examples include time to death, treatment failure or relapse; Leung et al. (1997) cited some further public health examples. Often it is not possible to observe all failures in the sample being studied, especially if the study terminates after a fixed follow-up period. In clinical trials, for example, patients who die within the follow-up period have a failure time recorded; however, all that is known for those still alive at the end of the trial is that their failure time is longer than their follow-up time. Such individuals are said to be censored. Even though their data are effectively incomplete, estimation of survival probabilities and comparative treatment effects can proceed under certain assumptions. A key assumption is that of independent or noninformative censoring, which means that censored individuals cannot constitute a particularly high or low risk subgroup; rather they should be representative in terms of their risk of failure. The goal of this article is to clearly demonstrate the role of this assumption and to show how its violation can lead to biased estimates of survival and covariate effects.

2.2. The Kaplan–Meier curve

The Kaplan–Meier estimator $\mathrm{KM}(t)$ estimates the probability that the time-to-event or time-to-failure T exceeds any given value t (Kaplan and Meier, 1958). It is typically plotted as a function of t over the range of times of interest and is a decreasing curve with value 1 at time zero and other values given by:

$$\mathrm{KM}(t) = \prod_{i:s_i<t} \left(1 - r_{s_i}\right), \tag{1}$$

where $\{s_1, s_2, \ldots\}$ are the observed failure times and r_s is the estimated hazard or risk of failure at time s, among all individuals at risk of failure at time s.

From expression (1), it is clear that the key to an unbiased Kaplan–Meier estimator is an unbiased set of estimators of the hazards r_s at the observed failure times. With censoring, some individuals may be lost to follow-up before a given failure time s, in which case we cannot observe the complete at-risk population at this time. In this situation, the survival analyst estimates the hazard of failure at time s by the observed failure rate among those at risk and still under observation at s. For this to be unbiased, the individuals at risk and still under observation at s must be representative of the population at risk at s. Equivalently, the individuals censored before s cannot be a selectively high or low risk subgroup. If high-risk individuals tend to be censored, then those remaining will constitute a selective, low-risk sample, fewer events than expected will occur, and the estimated hazard will underestimate the true hazard. This is a case of dependent censoring; the selective censoring effectively induces a correlation between the censoring and

failure times. From Eq. (1), it is clear that underestimating hazards will inflate the Kaplan–Meier curve and lead to overestimation of survival. The reverse will occur if low-risk individuals tend to be censored.

Dependent censoring will occur to some extent in practically any cost-to-event analysis (Hallstrom and Sullivan, 1997; Lin et al., 1997). The problem arises because individuals tend to accrue costs at different rates, with those in poorer health using more resources and costing more per unit time. Consequently, individuals censored with low costs will tend to be those accumulating costs slowly, who in turn will tend to be those with lower costs-to-event. In practice the correlation between cost at censoring and cost-to-event may not be so extreme as to cause noticeable bias. However, although this correlation is unobservable, its presence in a real application is evidenced by the example in the next section, which shows inflation of the Kaplan–Meier curve. In theory, unless the mapping from time t to cost accumulated by time t is one to one, some degree of bias is to be expected. This can happen even if there is independent censoring on the time scale.

Fig. 1 demonstrates the magnitude of the bias of the Kaplan–Meier method applied to costs when individuals accumulate costs at different rates. The figure represents a 5-year study, with continuous accrual during the follow-up period. Thus, failure times and censoring times are completely independent on the interval 0 to 5 years. Patients accrue costs at a rate of US$1 per month or US$10 per month, each with probability 0.5. Fig. 1a and b show that the Kaplan–Meier method provides an excellent estimate of survival on the time scale, but that the methodology applied to costs can lead to substantial overestimation. The degree of bias is a function of the amount of censoring and the heterogeneity of the cost accrual rates. For instance, if individuals accumulate costs at a rate of either US$1 or US$2 (rather than US$10) per month, then the Kaplan–Meier estimate of the cost-to-event distribution Fig. 1c shows only slight bias compared with Fig. 1b. This is a result of the fact that the correlation between costs at censoring and costs at failure is 0.56 in the example depicted by Fig. 1b and only 0.25 in Fig. 1c.

A key feature of the previous example is that the maximum censoring time (5 years) is at least as large as the maximum failure time. In other words, the follow-up period is sufficient to cover the entire range of possible failure times. In practice, this is not always the case. For instance, the example in Section 2.3 has just 7 years of cost data on ovarian cancer patients, who may well survive beyond this time, particularly if their disease is localized at diagnosis. Section 3 presents some alternatives to the Kaplan–Meier approach which are appropriate when follow-up is not sufficient to cover the entire range of lifetimes.

2.3. Cox regression

Suppose now that rather than estimating the distribution of costs to an event, the goal is actually to relate cost-at-event to covariates like health status or age; in

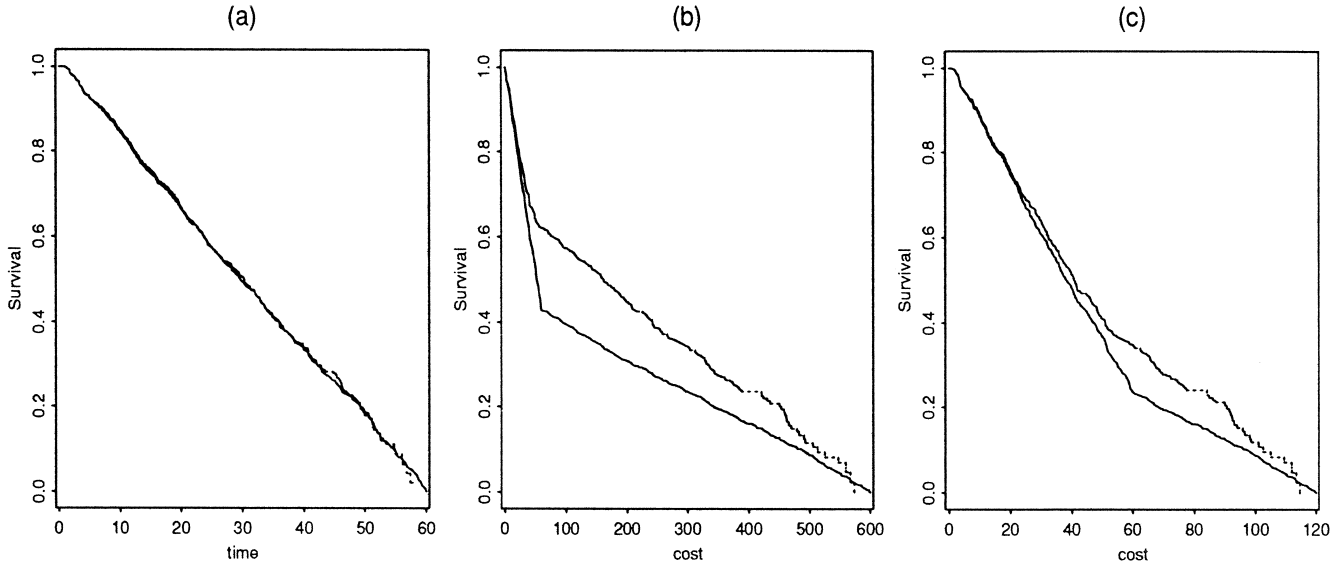

Fig. 1. Kaplan–Meier estimation in a cost-to-event problem with independent censoring in time. Simulation of 1000 subjects with survival times distributed uniformly between 1 and 60 months, and censoring times independent of survival and also distributed uniformly between 1 and 60 months. The solid curves are estimates based on complete data, i.e., without censoring. The dashed curves are the estimates under censoring. (a) Kaplan–Meier estimate of the survival distribution. (b) Kaplan–Meier estimate of the distribution of costs to event. Fifty percent of cases accrue costs at a rate of US$1 per month and the rest accrue costs at a rate of US$10 per month. (c) Kaplan–Meier estimate of the distribution of costs to event. Fifty percent of cases accrue costs at a rate of US$1 per month and the rest accrue costs at a rate of US$2 per month.

essence to perform a regression of costs on a collection of independent variables X. The most common regression approach used in time-to-event problems is Cox regression (Cox, 1972).

The Cox model is a description of the dependence of the risk of failure at any time t on the covariates X. It is semiparametric in that no assumptions are made about how the hazard rates vary with time; however, the hazards for different covariate values are assumed to be proportional with a ratio that is constant over time.

Since classical Cox regression relates the hazard at each time t to covariates, the model applied to costs relates the hazard at each cumulative cost c, to covariates. For illustration, consider a binary covariate X taking values 0 and 1. Suppose, for ease of discussion, that X is tumor stage at diagnosis in cancer patients; $X = 0$ is localized and $X = 1$ is metastatic disease. Suppose that the hazard for metastatic disease is a factor a times the hazard for localized disease. The hazard ratio a is termed the 'relative risk'. A relative risk of 2 in a cost analysis would mean that for metastatic cases, the hazard at any cost c, in terms of events per person-dollars at risk, is twice that for localized cases. This is not in itself a useful quantity, although it indirectly addresses the questions that are usually of interest in cost analyses so long as the Cox regression methodology is valid. These include the following: (1) Overall, how do the costs for localized and metastatic disease compare; (2) For a specific time-to-event, how do the costs compare, and (3) What is an estimate of the marginal cost difference between the two groups?

For Cox regression to be unbiased, independent censoring is required within groups formed by each level of the covariate X so that individuals still under observation are representative of the population at risk in each group, and observed events occur at the correct rate within each group. If censoring is dependent, the observed event rates in each group will be biased. If the dependent censoring mechanism is the same for all levels of X, then the estimate of the relative risk may still be unbiased; the errors caused by dependent censoring within each group may, in a sense, cancel out. However, if, for example, individuals at risk of failure are censored more often when $X = 1$, the observed failure rate for this level of X will be correspondingly lower and as a result, the relative risk a will be underestimated.

In practice, when using Cox regression for cost analysis, the accrual of costs at different rates leads to dependent censoring within subgroups defined by covariate levels. Covariates that affect the rate of cost accrual may lead to differential dependent censoring across groups. To demonstrate the bias that can arise when using Cox regression to analyze costs, we simulated a situation where for $X = 0$, survival is exponential with mean 20 months, and costs accrue at a rate of either US$1 or US$10 per month, each with probability 0.5. For $X = 1$, survival is exponential with mean 10 months, and costs accrue at a rate of either US$2 or US$20 per month, each with probability 0.5. This leads to a proportional hazards

model in costs, with a true relative risk of 1.0, since the increased rate of cost accrual is exactly balanced by the higher event rate when $X = 1$.

Assuming independent censoring in time with censoring times uniformly distributed on the range 0 to 20 months, the mean estimated relative risk over 100 simulations with 500 subjects per group is 1.2 with a standard deviation of 0.1. Thus, analysis of data from such a model would lead to the conclusion that the costs are lower for $X = 1$, which is not the case.

In this example, the different rates of cost accrual in the two groups imply differential dependent censoring in costs with independent censoring in time. A confirmation of this is the observation that the correlation between the cost at censoring and the cost at failure is higher in general for $X = 1$ than for $X = 0$. Consequently, the relative risk estimate is biased. In practice, the degree of bias will differ from one analysis to another, and will depend, among other things on the amount of censoring and the differential in survival and rates of cost accrual in the different groups. When comparing costs in two groups, bias will tend to be greater when the Kaplan–Meier estimate is biased only for one group than when the estimates for both groups are biased in the same direction. For example, bias will occur when rates of cost accrual are highly variable in one group and less so in the other.

Even if it is suspected that dependent censoring will not impact too severely on the bias of the estimated relative risk, the proportional hazards assumption will not in general be satisfied when costs are accruing at different rates. Consider a simple model where, for low X, costs accumulate at a rate of US$1 per month with probability p, or 10 per month with probability $1 - p$. Suppose that survival is exponential with mean m. Two natural models for the costs at high X might be either a different value of p or a different mean survival m, neither of which leads to proportional hazards in costs (see Appendix A). In the simulated example, the only way to achieve proportional hazards was to keep p constant at both levels of X, and to vary m and the rates of cost accrual together in a highly specific manner. We conclude that the appropriateness of the proportional hazards assumption would appear to be questionable in data sets where rates of cost accumulation vary among individuals.

3. Alternative approaches

In this section, we review a number of alternative approaches for estimating the distribution or mean of the costs to an event. We focus here on nonparametric approaches (Etzioni et al., 1996; Lin et al., 1997), since they are being proposed as alternatives to survival analysis, itself nonparametric. Parametric approaches to costs analysis have recognized limitations due to the frequent skewness of medical cost data and the presence of substantial numbers of observations with zero costs (Duan, 1983; Duan et al., 1983). Recent approaches to modeling survival simulta-

neously with a repeated measures process may have application to the modeling of medical costs over time, but these approaches tend to be parametrically based. Examples include modeling survival and CD4 counts (Wulfsohn and Tsiatis, 1997), as well as survival and hospitalization in AIDS patients (Lancaster and Intrator, 1998).

If the individual cost histories in terms of costs per unit time have been recorded, then an estimator of the mean cost-to-event is:

$$M_1 = \sum_i \hat{S}_i \hat{c}_i, \tag{2}$$

where $\hat{S}_i$ denotes the estimated survival probability and $\hat{c}_i$, is the average cost incurred among patients surviving to time period i. An alternative is:

$$M_2 = \sum_i \hat{s}_i \hat{C}_i, \tag{3}$$

where $\hat{s}_i$ is the estimated probability of the event occurring in time period i and $\hat{C}_i$ is the average cumulative cost among patients experiencing the event in time period i. This approach can be used if the cost histories have not been recorded. It has the disadvantage that incomplete cost histories, like those from censored individuals, cannot be used for estimation purposes unless further assumptions are made. However, in contrast to M_1, it is applicable if the distribution of the costs to an event is required, in which case the cumulative distribution function of the cost-to-event C is estimated by:

$$\hat{P}(C \le c) = \sum_i \hat{s}_i \hat{P}(C_i \le c). \tag{4}$$

In Eq. (4), $\hat{P}(C_i \le c)$ is the cumulative distribution function of the costs among individuals who experience the event in time period i.

When there is no censoring, M_1 and M_2 are identical; this can be easily seen by constructing a table in which the rows represent individuals and the columns the costs per unit time. M_1 is then obtained by constructing an appropriately weighted sum of the column totals and M_2 by a weighted sum of the row totals.

Both M_1 and M_2 have previously been used in the costs literature (Keeler et al., 1989; Hodgson, 1992; Riley et al., 1995). Unlike the Kaplan–Meier approach, they are appropriate when follow-up does not cover the entire range of failure times, in which case they have somewhat different interpretations. Suppose cost and survival data are only available for up to I months. Then, M_1 estimates the average cost of care for all individuals including survivors over I months, with individuals dying before I contributing their costs until death and individuals alive at I contributing their cumulative costs until this time. In contrast, M_2 estimates the average cost of care among individuals who die within I months. Similarly to M_2, expression (4) estimates the distribution function of the cumulative costs of

care among individuals who die within I months. [1] A version of M_2 that yields the same interpretation as M_1 is available in this case, and is given by:

$$M_3 = \sum_{i=1}^{I} \hat{s}_i \hat{C}_i + \hat{S}_I \hat{c}_I, \tag{5}$$

where $\hat{S}_I$ is the estimated probability of surviving beyond time period I, and $\hat{c}_I$ is the average cost until time I among individuals surviving beyond this time. Thus, like M_1, this version of M_2 requires the individual cost histories to be recorded. Lin et al. (1997) present a detailed simulation-based analysis of the relative performance of M_1 and M_3 under both light and moderate censoring; both estimators are seen to be far more accurate than the estimate based on the Kaplan–Meier approach.

Standard error expressions are available for the alternative estimators M_1, M_2, and M_3 (Lin et al., 1997). Lin et al. (1997) also prove that the estimators are asymptotically consistent and normally distributed; these results can be used for inference and are useful when comparisons of expected costs are required. Indeed, the results of Lin et al. (1997) allow mean costs at different levels of categorical covariates to be compared; this methodology therefore provides a valid alternative to the use of Cox regression in many cases.

Since both M_1, M_2 and M_3 use separate estimates of costs and survival, they provide the researcher with the opportunity to use different data sources to estimate each of these quantities. This may be of value, especially if costs have been recorded on a relatively small sample of patients or over a short period of calendar time, but survival is available for a much larger sample and a longer observation period, as in disease registry data. So long as both data sources are reasonably comparable, the survival estimates from the larger data source may be used (Etzioni et al., 1996; Ramsey et al., 1997).

The problems of modeling quality-adjusted survival distributions and estimating expected quality-adjusted life years (QALY's) are analogous to the cost modeling problems discussed in this article (Glasziou et al., 1990). Therefore, methods valid for estimating expected QALY's may be applied directly to the estimation of expected costs under censoring. In quality-adjusted survival analysis, each time period spent alive is associated with a utility, that is, a measure of quality that ranges between zero and one, with a value of one representing perfect health and zero being equivalent to death. A utility in quality-adjusted survival analysis is analogous to a cost per time period in cost analysis. An individual's quality-adjusted survival is the integral of his utility curve over his survival time. In discrete

[1] In this case, the failure probabilities $\hat{s}_i$ in expressions (3) and (4) should be replaced by $\hat{s}_i/(1-\hat{S}_I)$, i.e., the probability of failing in time period i conditional on failure during or before time period I.

time, this translates into a cumulative sum of utilities that is exactly analogous to the notion of a cumulative cost of care.

A recent addition to the literature on quality-adjusted survival (Zhao and Tsiatis, 1997) provides a useful alternative to cost distribution estimates based on Eq. (4). The new approach draws upon the literature on statistical analysis with missing data and yields an estimator of the quality-adjusted survival that provides a distributional extension of M_1. Thus, if the maximum follow-up time I is insufficient to cover the entire range of survival times, the estimator is interpretable as an estimate of the distribution of cumulative costs until death or the maximum follow-up, which ever comes first.

4. Example

To illustrate the differences between the Kaplan–Meier and alternative approaches to estimating the distribution of costs-to-event, we analyzed a subset of the ovarian cancer data, presented in Etzioni et al. (1996) and Lin et al. (1997). The original dataset consisted of monthly Medicare reimbursements from 1984

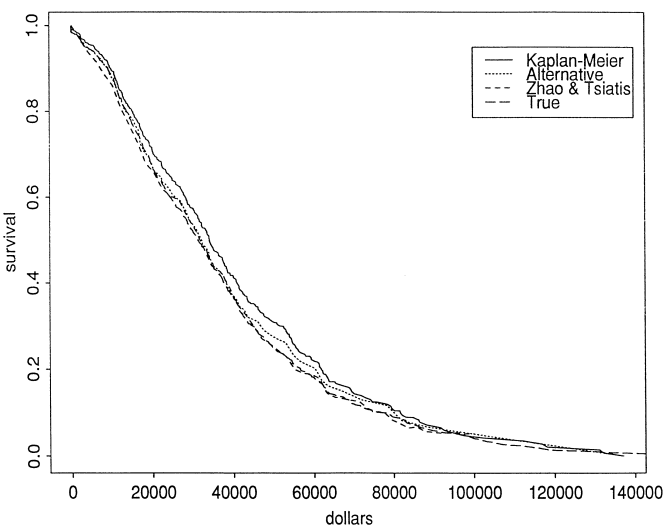

Fig. 2. Kaplan–Meier and two alternative estimators of the distribution of lifetime costs among 600 regional stage ovarian cancer patients diagnosed between 1984 and 1990 inclusive and dying within this time. An independent censoring distribution has been imposed on the survival times. The curve labelled 'Alternative' is based on expression (4); the curve labelled 'Zhao and Tsiatis' is based on Zhao and Tsiatis (1997).

through 1990 for 5012 Medicare beneficiaries diagnosed with ovarian cancer between 1973 and 1990 inclusive. The source of the data was the linked SEER–Medicare database (Potosky et al., 1993). In this analysis we used only the records on 600 regional stage ovarian cancer patients diagnosed on or after January 1984, and dying before the end of 1990, that is, those regional cases with complete cost information from the time of diagnosis. We imposed independent censoring on these individuals according to an exponential distribution with mean 24 months. This led to a sample with approximately 39% of individuals censored and a censoring distribution that covered the entire range of possible failure times. This latter condition was important to ensure the comparability of the various estimates, since they have different interpretations otherwise. We used the complete data to compute the true distribution function of lifetime costs which we then compared with the various estimates discussed in this article.

Fig. 2 plots the true distribution function of the lifetime costs of care together with the Kaplan–Meier estimate and the alternative estimates based on expression (4) and Zhao and Tsiatis (1997). The Kaplan–Meier estimate is seen to overestimate the true distribution of lifetime costs as expected, although the magnitude of the bias is not as severe as might be expected. This is probably due to the fact that

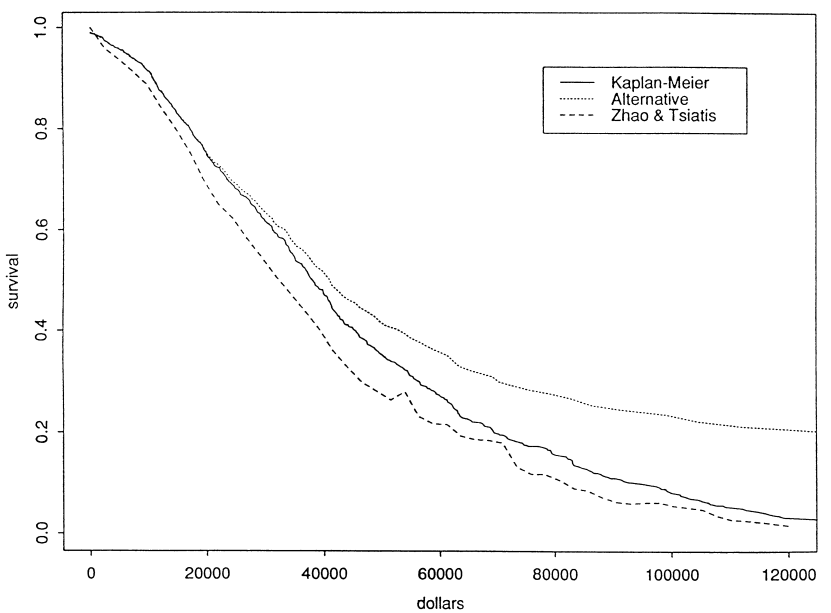

Fig. 3. Kaplan–Meier and two alternative estimators of the distribution of lifetime costs among 831 regional stage ovarian cancer patients diagnosed between 1984 and 1990 inclusive. The curve labelled 'Alternative' is based on expression (4); the curve labelled 'Zhao and Tsiatis' is based on Zhao and Tsiatis (1997).

exclusion of regional stage cases surviving beyond 7 years leaves a sample of cases with costs that are relatively homogeneous in terms of their rates of accrual over time. However, leaving the surviving cases in the sample would have led to estimators that are not in fact comparable. As expected, the alternative estimates show noticeably less bias than the Kaplan–Meier approach.

To examine the differences between the estimation approaches in a situation where survival could exceed the maximum follow-up, we analyzed the full set of 831 regional stage cases diagnosed on or after January 1984. Of these, 231 were alive at the end of follow-up. Fig. 3 shows that the different approaches yield quite different results in this case, but this is due more to the fact that they are estimating different quantities than to any bias considerations. The estimator labelled 'alternative' represents the distributional extension of M_2 and suggests higher costs than the others because it is indicative of the distribution of costs among individuals dying within 7 years. The Zhao and Tsiatis estimator estimates the distribution of 7-year costs, with individuals dying before 7 years contributing their costs until death, and those alive for the full 7 years their cumulative costs over this period. The Kaplan–Meier curve does not, as already noted, represent the distribution of lifetime costs, since the lifetimes are restricted to those deaths occurring within 7 years whereas no such restriction exists among censored observations. What the Kaplan–Meier curve does estimate in this case is difficult to say.

5. Discussion

The methodological issues that arise in the economic evaluation of medical care are many, but three issues have been a particular focus for analysts in recent years. The first is that medical care costs accrue unevenly over time for most individuals. In particular, time periods of months or years can pass between times at which health care costs are incurred. In databases that track health utilization over time for populations, this is manifested as a '0–1' problem: for most months, no health care costs accrue for the majority of the observed population. Second, in most situations, medical care costs for a given population are not normally distributed: a minority of individuals incur disproportionately high medical care costs compared to the rest of the population. Economists have traditionally relied on multi-part modeling and semiparametric techniques to address these issues (Duan, 1983; Duan et al., 1983). Finally, the issue of censoring has recently been raised in the context of economic analysis along side clinical trials, (Dudley et al., 1983; Quesenberry et al., 1989; Fenn et al., 1995) although in fact, problems created by censoring are also present in most retrospective analyses of population cost data (Medicare records, for example). Certainly, the analytical problems created by databases with a large mass of observations at zero, skewedness of cost distributions, and censoring have challenged those who wish to conduct hypothesis tests

of economic data in such settings as retrospective cost analyses and cost-effectiveness analyses conducted alongside clinical trials (Rutten-van Molken et al., 1994).

At first glance, survival analysis seems to be an appealing approach for addressing the issues raised above. Not only is survival analysis appropriate for censored data, but it also easily accommodates a variety of distributional forms. In this article, we have demonstrated the pitfalls of applying survival analysis techniques to cost estimation. The first issue is the real possibility of dependent censoring, which can inflate Kaplan–Meier-derived cost estimates and bias the coefficients in multivariate Cox regressions. A second problem is the difficulty of interpreting Kaplan–Meier-derived cost estimates when survival exceeds the maximal censoring time. Finally, the fundamental proportional hazards assumption may well be violated when Cox regression models are applied to medical cost data. All of these issues raise the possibility that cost estimates will be biased when survival analysis is used to analyze cost data. Although in practice, the bias may not be severe, assessing its magnitude is at best difficult, if not impossible. We therefore recommend against the use of these approaches in general.

We have outlined alternative ways to analyze costs when survival exceeds the maximum censoring time. All have the advantage that they can accommodate large masses of observations with zero costs and non-normal cost distributions. Each yields estimates of average costs and standard errors, thus permitting hypothesis testing of costs for alternative treatment arms in clinical trials. An analog of Cox regression for relating costs to covariates is not, to our knowledge available. This would seem to be an important area for future research. Finally, an important and under-appreciated application of empirically derived cost estimates from time-to-event studies is in estimating the numerator of the incremental cost-effectiveness (ICE) ratio. Traditional ICE ratios involve point estimates of costs according to various treatment groups, often derived from a nonstochastic source. These point estimates and resultant incremental ratios require variance estimates for analytic and hypothesis testing purpose (O'Brien et al., 1994). The techniques presented in this paper allow for variance estimates of the incremental cost values. Combined with variances on outcome data from traditional methods, and statistical methods for computing the variance of cost-effectiveness ratio estimates (Wakker and Klaassen, 1995; Siegel et al., 1996; Willan and O'Brien, 1996; Laska et al., 1997), an ICE ratio with confidence intervals can now be constructed for hypothesis testing.

Acknowledgements

Research supported in part by contract NCI N01-CN-05230. We thank Barbara KcKnight for useful discussion, Margaret Pepe, Nicole Urban and Wendy Leisenring for comments on an earlier draft, and Arnie Potosky for making the combined SEER–Medicare database available to us.

Appendix A

Illustration of the lack of validity of the proportional hazards assumption in a simple model with costs accruing at different rates.

Assume a binary covariate X with a value of 0 or 1. Suppose that given $X = x$, survival is exponential with mean $(1/\lambda_x)$ and costs accumulate at a rate of US$$a_x$ per month with probability p_x, and US$$b_x$ per month with probability $1 - p_x$.

We have the following model:

$$f_x(t) = \lambda_x \exp\{-\lambda_x t\};$$

where t is survival time and $f_x(t)$ is the probability density function of t given $X = x$. Denote an individual's cost-to-event by $c(t)$. Then, $c(t) = a_x t$ with probability p_x and $c(t) = b_x t$ with probability $1 - p_x$. The probability density function of $c(t)$ is

$$f_x(c(t)) = \frac{\lambda_x}{a_x}\exp\left\{\frac{-\lambda_x}{a_x}t\right\}p_x + \frac{\lambda_x}{b_x}\exp\left\{\frac{-\lambda_x}{b_x}t\right\}(1-p_x),$$

and the survivor function is

$$S_x(c(t)) = 1 - F_x(c(t)) = \exp\left\{\frac{-\lambda_x}{a_x}t\right\}p_x + \exp\left\{\frac{-\lambda_x}{b_x}t\right\}(1-p_x),$$

where $F_x(c(t))$ is the cumulative distribution function of $c(t)$.

The cost hazard given $X = x$ is the ratio of $f_x(c(t))$ to $S_x(c(t))$, that is:

$$h_x(c(t)) = \frac{\frac{\lambda_x}{a_x}\exp\left\{\frac{-\lambda_x}{a_x}t\right\}p_x + \frac{\lambda_x}{b_x}\exp\left\{\frac{-\lambda_x}{b_x}t\right\}(1-p_x)}{\exp\left\{\frac{-\lambda_x}{a_x}t\right\}p_x + \exp\left\{\frac{-\lambda_x}{b_x}t\right\}(1-p_x)}$$

and the ratio of the hazard when $X = 1$ to the hazard when $X = 0$ is:

$$\frac{h_1(c(t))}{h_0(c(t))} = \frac{\frac{\frac{\lambda_1}{a_1}\exp\left\{\frac{-\lambda_1}{a_1}t\right\}p_1 + \frac{\lambda_1}{b_1}\exp\left\{\frac{-\lambda_1}{b_1}t\right\}(1-p_1)}{\exp\left\{\frac{-\lambda_1}{a_1}t\right\}p_1 + \exp\left\{\frac{-\lambda_1}{b_1}t\right\}(1-p_1)}}{\frac{\frac{\lambda_0}{a_0}\exp\left\{\frac{-\lambda_0}{a_0}t\right\}p_0 + \frac{\lambda_0}{b_0}\exp\left\{\frac{-\lambda_0}{b_0}t\right\}(1-p_0)}{\exp\left\{\frac{-\lambda_0}{a_0}t\right\}p_0 + \exp\left\{\frac{-\lambda_0}{b_0}t\right\}(1-p_0)}}.$$

Therefore, for example, simply varying a_x, b_x, p_x or λ_x individually between the two groups will not lead to proportional hazards. Indeed, it is quite difficult to find

settings for a_x, b_x, λ_x and p_x that will lead to proportional hazards; the example in Section 2.3 varies λ_x, a_x and b_x so that $\lambda_0/a_0 = \lambda_1/a_1$ and $\lambda_0/b_0 = \lambda_1/b_1$ while keeping $p_1 = p_0$. To summarize, the heterogeneity of rate of cost accrual in the two groups leads to a costs variable that is a mixture of random variables. The resulting probability density functions and hazards are complex and are not necessarily proportional across groups with differing survival distributions or cost accrual rates.

References

Cox, D.R., 1972. Regression models and life-tables (with discussion). Journal of the Royal Statistical Society B 34, 187–220.

Duan, N., 1983. Smearing estimate: a nonparametric retransformation method. Journal of the American Statistical Association 78, 605–610.

Duan, N., Manning, W.G., Morris, C.N., Newhouse, J.P., 1983. A comparison of alternative models for the demand for medical care. Journal of Business and Economic Statistics 1, 115–126.

Dudley, R.A., Harrell, F.E., Smith, L.R., Mark, D.B., Califf, R.M., Pryor, D.B., Glower, D., Lipscomb, J., Hlatky, M., 1983. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. Journal of Clinical Epidemiology 46, 171–261.

Etzioni, R., Urban, N., Baker, M., 1996. Estimating the costs attributable to a disease with application to ovarian cancer. Journal of Clinical Epidemiology 49, 95–103.

Fenn, P., McGuire, A., Phillips, V., Backhouse, M., Jones, D., 1995. The analysis of censored treatment cost data in economic evaluation. Medical Care 33, 851–863.

Fenn, P., McGuire, A., Backhouse, M., Jones, D., 1996. Modelling programme costs in economic evaluation. Journal of Health Economics 15, 115–125.

Glasziou, P.P., Simes, R.J., Gelber, R.D., 1990. Quality adjusted survival analysis. Statistics in Medicine 9, 1259–1276.

Hallstrom, A.P., Sullivan, S.D., 1997. On estimating costs for economic evaluation in failure time studies. Medical Care, in press.

Heckman, J.J., Singer, D.D., 1985. Longitudinal Analysis of Labor Market Data. Cambridge UK Econometric Society Monographs (10). Cambridge Univ. Press, Cambridge, UK.

Hodgson, T.A., 1992. Cigarette smoking and lifetime medical expenditures. The Milbank Quarterly 70, 81–125.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457–481.

Keeler, E.B. et al., 1989. The external costs of a sedentary lifestyle. American Journal of Public Health 79, 975–981.

Kiefer, N.M., 1988. Economic duration data and hazard functions. Journal of Economic Literature 26, 646–679.

Lancaster, T., 1990. The Econometric Analysis of Transition Data. Econometric Society Monographs (17). Cambridge Univ. Press, Cambridge, UK.

Lancaster, T., Intrator, O., 1998. Panel data with survival: hospitalization of HIV-positive patients. Journal of the American Statistical Association 93, 46–53.

Laska, E.M., Meisner, M., Siegel, C., 1997. Statistical inference for cost-effectiveness ratios. Health Economics 6, 229–242.

Leung, K., Elashoff, R.M., Abdelmonem, A.A., 1997. Censoring issues in survival analysis. Annual Review of Public Health 18, 83–104.

Lin, D.Y., Feuer, E.J., Etzioni, R., Wax, Y., 1997. Estimating medical costs from incomplete follow-up data. Biometrics 53, 113–128.

O'Brien, B.J., Drummond, M.F., Labelle, R.J., 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Medical Care 32, 150–163.

Potosky, A.L., Riley, G.F., Lubitz, J.D., Mentnech, R.M., Kessler, L.G., 1993. Potential for cancer related health services research using a linked medicare–tumor registry database. Medical Care 31, 732–748.

Quesenberry, C.P., Fireman, B., Hiatt, R.A., Selby, J.V., 1989. A survival analysis of hospitalization among patients with Acquired Immunodeficiency Syndrome. American Journal of Public Health 79, 1643–1647.

Ramsey, S.D., Etzioni, R., Troxel, A., Urban, N., 1997. Optimizing sampling strategies for estimating Quality-Adjusted Life years. Medical Decision Making, in press.

Riley, G., Potosky, A.L., Lubitz, J.D., Kessler, L.G., 1995. Medicare payments from diagnosis to death for elderly cancer patients by stage at diagnosis. Medical Care 33, 828–841.

Rutten-van Molken, M.P., van Doorslaer, E.K., Van Vliet, R.C., 1994. Statistical analysis of cost outcomes in a randomized controlled clinical trial. Health Economics 3, 333–345.

Siegel, C., Laska, E.M., Meisner, M., 1996. Statistical methods for cost-effectiveness analysis. Controlled Clinical Trials 17, 387–406.

Wakker, P., Klaassen, M.P., 1995. Confidence intervals for cost/effectiveness ratios. Health Economics 4, 373–381.

Willan, A.R., O'Brien, R.J., 1996. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. Health Economics 5, 297–305.

Wulfsohn, M.S., Tsiatis, A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339.

Zhao, H., Tsiatis, A., 1997. A consistent estimator for the distribution of quality adjusted survival time. Biometrika 84, 339–348.