Working paper No: 13/2009 February 2009 LSE Health

Maria Raikou and Alistair McGuire

Parametric estimation of medical care costs under conditions of censoring

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Parametric estimation of medical care costs under conditions of censoring

Maria Raikou and Alistair McGuire

Working paper No. 13/2009

First Published in February 2009
LSE Health
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

British Library Cataloguing in Publication Data
A catalogue record for this publication is available from the British Library
ISBN [978-0-85328-006-4]

Corresponding author(s)
Dr Maria Raikou and Professor Alistair McGuire
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Email: m.raikou@lse.ac.uk or a.j.mcguire@lse.ac.uk

Abstract

This paper is concerned with a set of parametric estimators that attempt to provide consistent estimates of average medical care costs under conditions of censoring. The main finding is that incorporation of the inverse of the probability of an individual not being censored in the estimating equations is instrumental in deriving unbiased cost estimates. The success of the approach is dependent on the amount of available information on the cost history process. The value of this information increases as the degree of censoring increases.

Key words: Cost of medical care; Censoring; Survival analysis; Regression analysis; Health care economic evaluation.

JEL Classification: C000; C100

Content

1. Introduction

The necessity of adopting economic evaluation in the health care sector arises because the market fails to fulfil the conditions required to ensure an efficient allocation of resources. Economic evaluation provides a method for determining the point of efficiency; that is, the point at which the allocation of resources leads to maximisation of social welfare. In the process of achieving the optimal resource allocation, alternative states have to be evaluated, each one associated with different individual welfare levels. Given that any alternative state of resource allocation will normally result in an improvement in welfare for some individuals and a deterioration for others, interpersonal comparisons of utility have to be made in order to determine whether there is a net gain in social welfare. The choice becomes then either to consider situations in which unambiguous welfare improvements are possible or to consider a wider range of situations by making interpersonal comparisons. In the former case, evaluation of alternative states is undertaken based on the Pareto principle according to which welfare improvement occurs if resource allocation is such that an individual is made better off without making another individual worse off. In the latter case, value judgements must be made to determine whether there are net gains in welfare. In this context, cost-benefit analysis is implemented specifically as a means of achieving Pareto welfare. In the health care sector, where the monetary valuation of outcomes is complex, cost-effectiveness commonly replaces cost-benefit analysis as a method of identifying patterns of health care resource allocation. If relative valuations can be attached to health states, then cost-effectiveness may encompass cost-utility analysis. All these methods, cost-benefit, cost-effectiveness and cost-utility analysis, have specific problems of implementation which have long been discussed in the welfare economics literature. Most recently, three particular themes have come to dominate the literature in health economics.

First, there has been increasing consideration of the specific technical conditions under which cost-effectiveness and cost-utility analyses relate to cost-benefit analysis, whose objective is to identify Pareto optimal states consistent with the maximisation of social welfare (Garber, 2000; Weinstein and Stason, 1976; Weinstein and Zeckhauser, 1973). Secondly, there has been growing criticism of the traditional definition of welfare as based on Pareto optimality and utility maximisation (Williams and Cookson, 2000; Tsuchiya and Williams, 2001). It has been suggested that the definition of welfare ought to take account of concepts that are not solely utility based. The justification for this approach derives from Sen's argument that welfare is not only defined by means of utility but is also related to fundamental attributes, which he refers to as 'basic capabilities' (Sen, 1982). On this basis, proponents of the notion of extra-welfarism have suggested that efficiency may be defined with regard to the maximisation of health and not utility *per se*. As such, these capabilities may be related to cardinal measurements of health benefit, allowing the problems imposed by interpersonal comparisons to be overcome. Within this context, the role of economic evaluation is not to determine the optimal allocation of health care resources that will maximise utility-based welfare, but rather to supply the relevant decision makers with information that assists their assessment of the appropriate allocation of health care resources. Under this interpretation, cost-effectiveness and cost-utility analyses are not necessarily related to cost-benefit analysis as there is no attempt to follow Paretian notions of efficiency. Cost-effectiveness and cost-utility analyses become appropriate allocative tools in their own right. The third theme that has dominated the literature has dealt with measurement issues given that any economic evaluation involves measurement of the costs incurred by and the benefits derived from a health care intervention. Particular emphasis has been given to the measurement of the benefits derived from a given health outcome. At the same time, there is a sizeable literature that considers the measurement of the relative valuations of health states using non-monetary values (Dolan, 2000, 2001). A directly related literature considers the relative strengths and weaknesses of the various measures. Other measurement-related issues that have received some attention in the literature include adjustments for missing data, the measurement of indirect costs and the transferability of findings across different regulatory environments.

This paper relates to the general aspect of measurement issues. Having focused on measurement problems raised within the context of analysing health outcome data, the literature has generally given less attention to the issues that arise in the analysis of cost data. With respect to cost, the matter most commonly addressed is the definition and measurement of indirect costs (Sculpher, 2001). The measurement of direct costs has received less attention. There is a relatively small literature which considers the appropriate definition of direct costs and their relationship to opportunity costs and charges (Brouwer et al, 2001; Dranove, 1996). There is limited consideration, however, of the impact that different data collection methods and different methodological approaches employed in analysing cost data have on the estimates of cost statistics. The limited information available with regard to direct cost measurement in general and the lack of a well-established methodology in dealing with particular statistical issues arising in the analysis of treatment costs are themselves a justification for the subject matter that follows.

The study of medical care costs presents a number of analytical difficulties. An issue of particular concern arises when a specific form of incomplete information is present in the data, a condition which is referred to as censoring. An individual whose behaviour with respect to the variable under study is not observed for the full duration of interest is said to be censored. Thus patients who are lost to follow-up, drop out of the study or are observed until the termination of the study period without having reached the event of interest will be right censored.[1] Given the bias imparted by this loss of information, estimators of statistics of interest must account for the presence of censoring in the data. While both parametric and non-parametric approaches have been applied to the analysis of time-to-event data, yielding estimators which successfully account for censoring, applications where the random variable of interest is cost-to-event have generally placed emphasis on non-parametric estimation (Lin et al, 1997; Bang and Tsiatis, 2000).

The primary advantage of non-parametric models is that they are free of assumptions concerning the distribution of the variable of interest. There are circumstances, however, where parametric methods may be the preferred or necessary alternative. A parametric approach can provide information on the pattern of cost accumulation by assessing individual covariate effects on cost or by modelling the relationship between cumulative cost and time. As such, parametric models can provide an instrument for extrapolating estimates of costs over the study period to different populations or to points in time exceeding the duration of the study. Although the usefulness of parametric modelling in analysing censored cost data has been acknowledged in the literature and there have been a number of approaches introduced recently, their validation has not been explored adequately. This paper presents an attempt to address this issue by studying a set of alternative parametric estimators of cost under conditions of censoring and by assessing their performance empirically under conditions of heavy censoring.

As in the case of non-parametric models (e.g. Fenn et al, 1995), the earliest attempts to account for censoring in deriving estimates of mean cost using a parametric approach involved direct application of the classical survival techniques to censored cost data. The Cox proportional hazards model and the Weibull and exponential models were applied, for example, by Dudley et al (1993) and Fenn et al (1996) in studying covariate effects on cumulative cost and in providing mean cost estimates over the study period. These approaches, however, generally lead to biased estimates for the same reason as their non-parametric counterparts; that is, due to the presence of dependent censoring between the variable of interest and its censoring variable (Lin et al, 1997; Etzioni et al, 1999). As in the non-parametric approach to the analysis of time-to-event data, the central concept in the semiparametric and parametric approaches is the conditional probability of an event occurring at a given point in time, given that it has not occurred until that point in time as modelled through the hazard functions. For all these models, independent censoring requires that individuals who are censored at time *t* (after allowing for covariates) are representative of all individuals who

[1] Another form of censoring, referred to as left censoring, is associated with incomplete information due to individuals entering the study at different points of progression to endpoint, but it is relatively uncommon in medical studies and is not considered in this paper.

are still under observation at *t*. When applying these approaches to modelling cost-to-event data, individuals who are censored having attained a particular cost level must be representative of all individuals who are still under observation having attained that cost level. This is not normally the case, as patients who are in poorer health states generate higher costs per unit of time and consequently are expected to generate higher cumulative costs at both the failure time and the censoring time, thus implying positive correlation between cost at failure and cost at censoring. Failure of these approaches to account for censoring in the cost estimates has led to two proposed alternatives. The first adopts a regression approach where cost is modelled as a function of failure time and adjustment for censoring in the cost estimates is achieved through adjusting failure time for censoring. The second uses a linear regression approach where cost is related to a set of individual covariates and adjustment for censoring in the cost estimates is performed by using the inverse of the probability of an individual not being censored in the estimating equations. These estimators of cost together with their properties and underlying assumptions are considered below.

Assessment of the estimators' performance is achieved through direct comparison of the resultant parametric estimates to their non-parametric counterparts derived from the application of a set of previously studied (non-parametric) estimators, using the same dataset as presented in Raikou and McGuire (2004). This allows exploration of whether the estimators' asymptotic properties are maintained in a practical setting. The paper proceeds as follows. The general setting for the analysis is outlined first and the set of parametric estimators for cost together with the assumptions underlying their validity are then presented. The resultant cost estimates derived from the application of the alternative regression methodologies to a medical dataset which exhibits heavy censoring follow. Some concluding remarks and suggestions are then given.

2. Analytical framework

2.1. General setting

The basic aim of the approaches presented below is to derive an estimate of the mean total cost $\mu = E(M)$ and its variance over a specified period when the data is right censored, where the random variable *M* denotes the total cost for a patient during some specified time *T* and *E* denotes expectation. The distribution of the random variable *T* is assumed continuous over $(0, L]$, where *L* denotes the upper bound of *T* and *M* is the total cost incurred by a patient up to a maximum of *L* units of time. To accommodate censoring, a potential time to censoring denoted by *U* is defined and letting *T* denote time to death, the observables from a study in the presence of censoring are $X = \min(T, U)$, i.e. the last contact date; $\delta = I(T \le U)$, where $I(\cdot)$ is the indicator function taking the value of 1 when the argument is true (i.e. if the observation is uncensored) and zero otherwise; the cost accrued up to time *X* and other intermediate cost history for each subject, i.e. $M^H(t) = \{M(u), u \le t\}$, where $M^H(t)$ denotes the cost history up to time *t*, $M = M(T)$, with $M(u)$ being the known accumulated cost up to time *u* and *u* denoting points in time at which cost information becomes available. Letting $Z = (Z_1, ..., Z_p)'$ denote a $p \times 1$ vector of the covariates of interest, the observable data for *n* individuals are then the independent and identically distributed random vectors
$\{X_i = \min(T_i, U_i),\ \ \delta_i = I(T_i \le U_i),\ \ M_i^H(X_i), Z_i\}, i = 1, ..., n$, where *i* identifies an individual.

2.2. Least squares regression analysis with randomly right-censored data

Assuming the general setting as defined above and defining $T^* = \min(T, L)$, with Z being a $p \times 1$ vector of covariates whose effect on the cumulative cost at T^* one wishes to study, the methodology presented in this section introduced by Lin (2000) attempts to adjust the estimates

derived by the linear model given as $M = \beta' Z + \varepsilon$, where β is a $p \times 1$ vector of unknown regression parameters and ε is a zero-mean error term with an unspecified distribution for censoring. The first term of *Z* is set equal to 1 so that the first term of β corresponds to the intercept. In the absence of censoring, β is estimated by the least-squares normal equation

$$\sum_{i=1}^{n} (M_i - \beta' Z_i) Z_i = 0$$

In the presence of censoring, estimation by the above equation will lead to biased estimates for the regression parameters (Lancaster, 1990; Green, 1997). A naïve approach is to estimate the model by including only the uncensored cases in the estimation process. The regression parameters are again estimated by the least-squares normal equation but now only individuals with complete cost observations contribute information to the estimation process. As is the case in any similar missing data situation, such an analysis, referred to as complete case analysis, which totally discards the cases with missing values, leads to loss of information which could be a substantial problem if the proportion of cases with missing values is high. On this basis, the approach has been deemed useful only for providing a baseline method for comparisons. In contrast, the approach proposed by Lin (2000) accounts for the presence of censoring as follows. Under the assumption of a continuous distribution for failure time over $(0, L]$ and a continuous distribution of censoring time with censoring arising in a completely random manner, time to censoring has the survivor function $K(u) = pr(U > u)$, i.e. the survivor function $K(u)$ evaluated at a point in time *u* gives the probability of an individual not being censored at *u*. Defining $\delta_i^* = I(U \geq T^*)$ under random censoring conditions, the estimating equation for β is modified as:

$$\sum_{i=1}^{n} \frac{\delta_i^*}{K(T_i^*)} (M_i - \beta' Z_i) Z_i = 0$$

which implies that only individuals with complete cost observations over the duration of interest contribute cost information to the estimation process. The unknown survivor function $K(\cdot)$ is estimated by the Kaplan-Meier estimator (Kaplan and Meier, 1958) based on the data $\{X_i = \min(T_i, U_i), 1 - \delta_i, i = 1, ..., n\}$ as $\hat{K}(t) = \prod_{u \leq t} \left\{ 1 - \frac{dN^c(u)}{Y(u)} \right\}$, where $N^c(u) = \sum_{i=1}^{n} N_i^c(u)$ counts the number of individuals censored over time with $N_i^c(u) = I(X_i \leq u, \delta_i = 0)$ and $Y(u) = \sum_{i=1}^{n} Y_i(u)$ counts the number of individuals at risk over time with $Y_i(u) = I(X_i \geq u)$. Replacing the survivor function $K(\cdot)$ with its consistent Kaplan-Meier estimator results in the following estimating equation for β:

$\sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*)} (M_i - \beta' Z_i) Z_i = 0$, whose solution is given as

$$\hat{\beta} = \left\{ \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*)} Z_i^{\otimes 2} \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*)} M_i Z_i \text{ , where } \alpha^{\otimes 0} = 1, \alpha^{\otimes 1} = \alpha, \alpha^{\otimes 2} = \alpha\alpha' .$$

Thus the main idea underlying this approach is to weight the uncensored observations by the inverse of the probability of an individual not being censored evaluated at the time of the individual's failure. The idea underlying the use of this specific weight is that under conditions of

independent censoring, at time T_i, $K(T_i) = pr(U > T_i)$ is the probability that individual *i* has survived to T_i without being censored. Therefore any individual who is observed to die at T_i represents on average $1/K(T_i)$ individuals who might have been observed if there was no censoring. The same idea underlies the approach by Koul et al (1981) within the context of failure time regression analysis when the dependent variable (time to event) is subject to censoring. Lin (2000) studies the asymptotic properties of this estimator and derives estimates for its covariance matrix for large samples using the martingale version of the central limit theorem. The mean cost over $(0, L]$ can then be estimated as $\hat{M} = \hat{\beta}'\tilde{Z}$, where $\tilde{Z}$ denotes the covariates vector evaluated at the mean values of the covariates.

2.3. Least squares regression analysis with randomly right-censored data: multiple time intervals

The second approach presented by Lin (2000) extends the previous idea in situations where information on individual cost histories is available at various points in time over the duration of interest. The main purpose of this method is to increase efficiency by allowing use of cost information not being used by the preceding estimator. Adopting the same framework as Liang et al (1986), Lin (2000) models the marginal expectation of cost at each point in time for which cost information is available as a function of the covariates as follows. The duration of analysis $(0, L]$ is partitioned into *K* subintervals $(t_{k-1}, t_k], (k = 1,..., K)$, with $t_0 = 0$ and $t_K = L$, and for each subinterval *k* the following linear model is assumed:

$$M_{ki} = \beta_k' Z_i + \varepsilon_{ki} \qquad k = 1,..., K \qquad i = 1,..., n$$

where for individual *i*, $M_{ik} = M_i(t_k) - M_i(t_{k-1})$ is the cost incurred over subinterval $(t_{k-1}, t_k]$, β_k $(k = 1,..., K)$ are $p \times 1$ vectors of unknown regression parameters and the error terms ε_{ki}'s are assumed to be independent among different subjects but allowed to be correlated within the same subject. By summing over all *k* subintervals, the linear model for the cost over the whole duration of interest becomes $M_i = \beta' Z_i + \varepsilon_i \qquad i = 1,..., n$, where $M_i = \sum_{k=1}^{K} M_{ki}$, $\beta = \sum_{k=1}^{K} \beta_k$, and $\varepsilon_i = \sum_{k=1}^{K} \varepsilon_{ki}$. Defining $T_{ki}^* = \min(T_i, t_k)$ and $\delta_{ki}^* = I(U_i \geq T_{ki}^*)$, i.e. $\delta_{ki}^* = I\{\min(T_i, t_k) \leq U_i\}$, the estimating equation for β_k $(k = 1,..., K)$ is given as $\sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{K}(T_{ki}^*)}(M_{ki} - \beta_k' Z_i) Z_i = 0$, where $\hat{K}(T_{ki}^*)$ is the Kaplan-Meier estimator for the probability of not being censored based on the dataset $\{X_{ki}, \delta_{ki}^*, i = 1,..., n\}$, where $X_{ki} = \min(T_{ki}^*, U_i)$. The solution to the above estimating equation is then given as:

$$\hat{\beta}_k = \left\{ \sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{K}(T_{ki}^*)} Z_i^{\otimes 2} \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{K}(T_{ki}^*)} M_{ki} Z_i \qquad \text{with}$$

$$\hat{\beta} = \sum_{k=1}^{K} \left[\left\{ \sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{K}(T_{ki}^*)} Z_i^{\otimes 2} \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_{ki}^*}{\hat{K}(T_{ki}^*)} M_{ki} Z_i \right].$$

Comparing this estimator with its counterpart from the previous approach, the gain in cost information is due to the fact that here a subject contributes cost information to the estimating

equations over all time intervals for which the individual is not censored, i.e. over all *k*s for which $U_i > \min(T_i^*, t_k)$. By contrast, in the previous estimator an individual only contributes cost information to the estimates if the individual has reached the event of interest or the individual's censoring time exceeds the maximum observed time in the study. In studying the asymptotic properties of this estimator, the same methodology as for the previous estimator is adopted and a consistent estimator for the covariance matrix is derived. The mean cost over $(0, L]$ can be estimated as $\hat{M} = \hat{\beta}'\tilde{Z}$, where $\tilde{Z}$ denotes the covariates vector evaluated at the mean values of the covariates.[2]

2.4. Two-stage regression

Carides et al (2000) proposed an estimator for mean cost in which the total cumulative cost is modelled as a function of failure time. Their method was introduced as an attempt to overcome the limitation of the Lin et al (1997) non-parametric approach associated with the requirement of a discrete censoring pattern to ensure the estimator's consistency. Their estimator is referred to as a two-stage estimator because at the first stage of the estimation process the expected cost at any given point in time is estimated as a function of failure time and at the second stage the estimated expected costs at given points in time are weighted by the Kaplan–Meier probability of death at these points in time. The estimate of mean total cost is derived as the sum over time of these weighted individual cost estimates. Under this model the mean cost is therefore given by:

$$\mu = \int_0^\infty g(t)\left|dS(t)\right|$$

where $g(t) = E(M|T = t)$ is the expected cost of an individual with survival time *T* and $S(t) = pr(T \geq t)$. The first stage involves deriving an estimator $\hat{g}(t)$ for $g(t) = E(M|T = t)$ using a regression approach. The authors suggest that the regression be performed only on the uncensored observations on the basis that the treatment costs of censored individuals typically differ from the treatment costs of uncensored individuals at the same point in time and inclusion of censored observations will therefore impart bias into the estimate of $g(t)$. The second stage of the estimation process involves the weighting of the estimated regression function $\hat{g}(t)$ by the Kaplan-Meier estimate of the probability of death at time *t*. The two-stage estimator of the mean cost over $(0, L]$ is then given as $\hat{\mu}_{TS} = \int_0^L \hat{g}(t)\left|d\hat{S}(t)\right|$

where $\hat{g}(t)$ is an estimator for $g(t) = E(M|T = t)$ and $\hat{S}(t)$ is the Kaplan-Meier estimator for $S(t) = pr(T \geq t)$, that is, $\hat{S}(t) = \prod_{s \leq t}\left\{1 - \frac{\Delta N(s)}{Y(s)}\right\}$, where $N(s) = \sum_{i=1}^{n} N_i(s)$ counts the number of individuals dying over time with $N_i(s) = I(X_i \leq s, \delta_i = 1)$ and $Y(s)$ counts the number of

[2] Both approaches described above are generalised to the case of covariate-dependent censoring. To accommodate covariate dependent censoring, Lin (2000) proposes using the proportional hazards specification (Cox, 1972) to model the effect of covariates on the censoring distribution allowing formulation of the dependence of censoring both on discrete covariates, which might be used as stratification variables, and on continuous covariates. The asymptotic properties for both these estimators and the expressions for the limiting covariance matrices reported by Lin (2000) are derived adopting the same analytical framework as for the case of covariate independent censoring and follow as a direct generalisation of the results presented above for the covariate independent censoring case. In addition, the methods are not restricted by the censoring pattern or by the number of covariates.

individuals at risk over time as defined above[3]. The choice of the functional form for $g(t)$ depends mainly on the data under consideration and the authors suggest use of either a parametric regression model or a non-parametric smoother. In the case of a parametric regression, the authors consider models which are, with or without some transformation of the data, linear in the coefficients, thus allowing use of the ordinary least squares regression technique to derive estimates for the regression parameters.[4]

Due to the consistency of the Kaplan-Meier estimator, consistency of the two-stage estimator is ensured if the parametric model $g(t)$ is consistently estimated. Although under specific parametric assumptions the two-stage estimator is asymptotically normal with variance estimator directly following from the specific statistical distribution, the authors recommend that for practical purposes the bootstrap method be used to derive standard error estimates for the mean, as they argue that the assumption of asymptotic normality is unlikely to be valid in most applications. The authors conclude that such a regression based approach, where the relationship between cost and failure time is specified through a parametric model, is advantageous compared to a non-parametric approach due to efficiency gains resulting from the use of such a relationship. On the other hand, this is only going to be the case if the parameterisation reflects the true functional form of cost and failure time. In the event of model misspecification, a non-parametric approach for estimating the relationship between cost and failure time will be preferred.

---

[3] If the last observed time corresponds to censoring in which case the Kaplan-Meier estimator is undefined (Kaplan and Meier, 1958), to ensure consistency the estimator can be expressed as $\hat{\mu}_{TS} = \int_0^L \hat{g}(t)\left|d\hat{S}(t)\right| + \overline{M}_{u \geq L}\hat{S}(L)$ where $\overline{M}_{u \geq L}$ is an estimate of cost accumulated over $(0, L]$ for patients who survive beyond *L*.

[4] The models considered in this analysis follow the authors' suggestions and are specified as:

1. Linear relationship between total costs and failure time as $M_i = \beta_0 + \beta_1 T_i + \varepsilon_i$, where the error terms are normally distributed with zero mean and finite variance, so that the two-stage estimator for mean cost is $\hat{M} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_t$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated from ordinary least squares regression using only the uncensored cost observations and $\hat{\mu}_t = \int_0^t \hat{S}(u)du$ is the Kaplan-Meier estimator for mean survival time over $(0, t]$.

2. Linear relationship between costs transformed on the natural logarithm scale and failure time as $\ln M_i = \beta_0 + \beta_1 T_i + \varepsilon_i$, where the error term has a lognormal distribution. The mean cost is $\hat{M} = e^{\hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_t}$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates from the ordinary least squares regression using the uncensored observations only and $\hat{\mu}_t$ is the Kaplan-Meier estimate for mean survival time. Given that the error distribution in the untransformed scale is unknown, Duan (1983) suggested using a non-parametric estimator for the untransformed scale expectation referred to as the smearing estimator. The estimate for the mean cost incorporating Duan's smearing estimator is $\hat{M} = e^{\hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_t} \frac{1}{n}\sum_{i=1}^{n} e^{\hat{\varepsilon}_i}$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates from the ordinary least squares regression on the uncensored observations only, $\hat{\mu}_t$ is the Kaplan-Meier estimate for mean survival time and $\hat{\varepsilon}_i$ are the ordinary least squares residuals.

3. Linear relationship between costs transformed on the natural logarithm scale once more and failure time as $\ln(\ln M_i) = \beta_0 + \beta_1 \ln T_i + \varepsilon_i$. The mean cost without smearing is $\hat{M} = e^{e^{\hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_t}}$ and with smearing is $\hat{M} = \frac{1}{n}\sum_{i=1}^{n} e^{e^{(\hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_t)} e^{\hat{\varepsilon}_i}}$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates from the ordinary least squares regression using the uncensored observations only, $\hat{\mu}_t$ is the Kaplan-Meier estimate for mean survival time and $\hat{\varepsilon}_i$ are the ordinary least squares residuals.

3. Methods and results

3.1. Methods

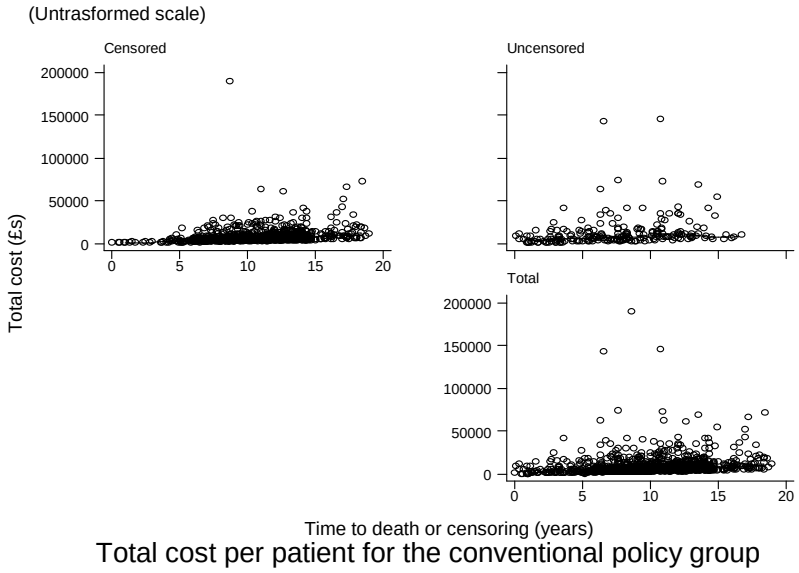
The parametric estimators defined above were applied to a medical dataset which exhibited extreme levels of censoring. The data were taken from a randomised controlled clinical trial and relate to a type 2 diabetic population of 3867 individuals allocated either to conventional policy (1138) or intensive policy (2729) with the aim of assessing the effectiveness of improved blood glucose control over a median follow-up period to death, the last date at which clinical status was known, or to the end of the trial period of 10 years. For each individual in the study, information on both clinical effectiveness and resource use was collected within the trial. Unit costs of resource use were attached to the recorded volume of resources to calculate the total cost per patient per year directly from the trial data and these were then aggregated to give a total cost per patient for the whole trial period. The analysis in this paper aims at deriving an estimate of average total cost over the trial period adjusting for censoring where an observation was defined as censored if the patient was not observed for the full time to death. Thus the failure event was all-cause mortality, resulting in 925 censored patients [81.3% censoring] and 213 failures in the conventional group and 2240 censored patients [82% censoring] and 489 failures in the intensive group by the end of the trial. Average follow-up time was equal to 9.9 years reaching a maximum of 18.93 years for the conventional group and 10.01 years reaching a maximum of 19.46 years for the intensive group. Despite the long duration of the trial, loss to follow-up and drop-out rates were negligible. The levels of censoring arising in the trial largely reflect the low mortality rates in both arms at the termination of the study. The assumption of independent censoring is valid in these data as censoring was not related to any cost or medical reasons.

All estimators were applied to these trial data within each randomisation group, where for each individual the observables were time to death or last contact, a variable taking the values of 0 or 1 indicating censoring or failure respectively, the annual costs and the total cost from the start of follow-up to death or the last contact date and a set of time independent covariates that represented measurements obtained on each individual at the start of the study on age, body mass index (bmi), fasting plasma glucose level (fpg), race and sex. The descriptive statistics for each of the covariates are shown in Table 1. As can be seen from Table 1, there are no differences in the baseline covariate values between the two groups. These covariates were deemed clinically meaningful given that fasting plasma glucose level provides the means of defining diabetes and body mass index gives an indication of obesity which is highly positively correlated with the risk of diabetes as is age. There is also evidence of racial differences in the incidence and prevalence of diabetes with, for example, higher rates in the Asian population. The fact that these covariates were deemed important explanatory variables for diabetes progression and complications does not imply that they will necessarily explain cost, especially as they were only measured at the start of the study. However, this represents the most frequent pattern of covariate measurements within a clinical trial setting, where interest lies in recording disease-specific predictive factors at the time of the individual's entry to the study and in certain cases at various points in time over the follow-up period.

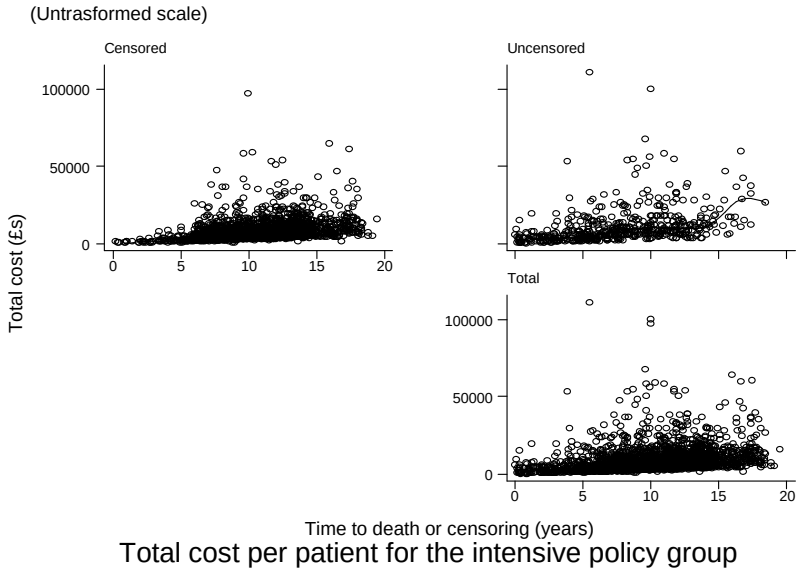
Table 1. Baseline covariate values in conventional and intensive policy groups

	Mean	Standard deviation	Minimum	Maximum
Conventional (*n*=1138)				
Age (years)	53.40	8.69	25.62	72
Bmi (kg/m^2)	27.80	5.46	17.57	55.68
Fpg (mmol/l)	8.48	2.03	5.5	17.5
Race	1.32	0.72	1	5
Sex	1.38	0.49	1	2
Intensive (*n*=2729)				
Age (years)	53.21	8.62	24.69	72
Bmi (kg/m^2)	27.49	5.07	16.59	60.61
Fpg (mmol/l)	8.61	2.14	5.4	19.9
Race	1.31	0.70	1	5
Sex	1.39	0.49	1	2

With respect to the methodology proposed by Carides et al (2000), the estimates of mean survival time used in all parameterisations were 15.65 years (*se*=0.21) for the conventional policy group and 15.96 years (*se*=0.18) for the intensive policy group. The analysis undertaken here derived mean estimates with and without smearing. An indication of the underlying relationship between treatment cost and study time is given in Figure 1 which plots the observed costs against time for the censored and uncensored populations for both trial arms.



Total cost per patient for the conventional policy group



Total cost per patient for the intensive policy group

Figure 1. Total cost per patient over the study period for conventional and intensive policy groups

The regression methodology proposed by Lin (2000) was applied to the data both when individual costs were available at the last contact date or death (failure time) and when multiple observations at different points in time were available for each individual. In the second model, annual time intervals were assumed for each individual because intermediate cost history was available for each subject on an annual basis. The classical linear regression model was estimated using the uncensored cases only as a baseline means for comparison to the alternative linear regression methodologies. All regression models aside from those proposed by Carides et al, which used failure time as the independent variable, were based on the set of covariates described above. Estimates of the variance associated with the mean estimators resulting from the above models were derived using the bootstrap approach with the estimates being obtained from 1000 replications. The reason for using the bootstrap approach is that the asymptotic variance estimators for the mean cost have not been defined. For compatibility purposes, estimates of the regression parameter standard errors for the Lin regression models as well as for the naïve ordinary least squares regression were also derived using the bootstrap approach.

3.2. Results

The results derived from the parametric approaches are shown in Table 3, while Table 2 reports the best non-parametric estimates obtained from a previous study by Raikou and McGuire (2004) as a means of assessing the parametric estimators' performance. Based on the conclusions drawn in this earlier study, two non-parametric estimators of within study average cost, one proposed by Lin et al (1997) and one proposed by Bang and Tsiatis (2000), both of which make use of information on intermediate cost histories, were deemed to perform adequately across a wide range of censoring levels. Given that these two estimators remained stable even under the extreme censoring conditions arising in the trial data, it can be confidently asserted that the resultant cost estimates reflect the true cost values.

Table 2. Best non-parametric estimators of mean cost

	Conventional		**Intensive**	
Estimator	Mean	Standard error	Mean	Standard error
Lin1 (Lin et al 1997) ($\hat{\mu}_{LIN1}$)	14006.2	897.73	13172	340.55
Bang and Tsiatis Partitioned ($\hat{\mu}_{P}$)	14639.48	1219.4	13839.67	445.6

Table 3. Parametric estimators of mean cost

	Conventional		**Intensive**	
Estimator	Mean	Standard error	Mean	Standard error
Carides et al regression models				
total cost on time	20353.71	2551.99	19548.07	1228.00
ln(total cost) on time without smearing	18086.73	2599.06	21096	1927.38
ln(total cost) on time with smearing	16070.78	1914.10	17939.50	1368.74
ln(ln(total cost)) on time without smearing	19080.38	3155.30	23132.24	2680.18
ln(ln(total cost) on time with smearing	18959.67	3152.28	21626.47	2545.91
Lin regression methodology				
Complete costs	14015.82	3588.94	17573.79	1961.70
Multiple intervals	14941.14	1274.07	13789.33	452.70
Naïve OLS	11708.78	1268.10	10845.21	693.58

With respect to the Carides et al two-stage estimator, the resultant mean cost estimates are high relative to the non-parametric estimates for both groups.[5] Moreover, the difference in average cost between the treatment arms is generally of the wrong expected direction given that the expected direction of difference in the average cost between the two trial arms is that the conventional group has higher average treatment costs than the intensive group, largely due to a higher hospitalisation rate. Although the approach is appealing as it attempts to model the time pattern of cost and is not restricted by assumptions concerning the censoring distribution, the analysis reveals the estimator's inadequate performance under all the parameterisations considered. This finding holds even when smearing estimates were obtained following a logarithmic transformation to account for positive skewness in the cost data. While model misspecification is liable to be a contributory factor, the estimator's inadequate performance is more likely attributed to the high degree of censoring present in the data. As the regression parameters are estimated using information from the uncensored cases alone, which in this case amounts to a mere 18% of the total number of observations and will reflect the bias imparted from a complete case analysis, it is to be expected that the estimated coefficients will not reflect the true parameter values, even assuming that the relationship between cost and failure time is correctly specified.

This postulate is supported further by the results obtained when the expected costs were estimated by a non-parametric regression approach. Carides et al (2000) recommend use of such a regression when there is not enough confidence in a specific parametric relationship between cost and survival time. The method adopted provides smoothed estimates of cost using locally weighted scatterplot smoothing (lowess) according to which the smoothed values of the dependent variable are derived by running a regression of the dependent variable on the independent variable using for each estimate the data at the estimation point and a small amount of data near the point. In lowess the regression is weighted so that the central point each time receives the highest weight and points farther away receive less. A separate weighted regression is estimated for each point in the data in order to provide the smoothed estimates. Applying this approach resulted in estimates of mean cost of 5674.92 (*se*=853.24) for the conventional group and 9407.87 (*se*= 3230.63) for the intensive group where the standard errors were obtained from 1000 bootstrap replications. Such an approach for deriving expected cost estimates, being free of assumptions about the functional form between cost and failure time, gives a strong indication that an equally important, if not more important, source of bias aside from model misspecification in the Carides et al estimator is the high level of censoring. This was to be expected based on the results obtained from the non-parametric estimators which only used cost information from the complete cases as presented in Raikou and McGuire (2004). Both the respective Lin et al (1997) and Bang and Tsiatis (2000) non-parametric estimators performed inadequately when only complete costs were included in the estimation process and both techniques showed dramatic improvement when information was increased by incorporating individual cost histories into the estimating equations.

Before considering the set of parametric estimators proposed by Lin (2000), the estimates derived from the naïve ordinary least squares regression are discussed.[6] The estimates derived from this approach are known to be biased as they are based on a complete-case analysis which ignores all censored observations, but as stated above, they provide a means for baseline comparisons to the alternative linear regression methodologies and in particular to Lin's (2000) regression models, which are a direct extension of this approach and use the same set of covariates. Although the naïve least squares regression resulted in the expected direction of the difference between the two arms of the trial, with the conventional group incurring higher costs on average than the intensive, the estimates of mean cost are low for both groups. This was anticipated as the information from censored observations is not used in the estimation process and it is known that the bias increases as the level of censoring increases. Comparison of the ordinary least squares cost estimates with the non-parametric uncensored cases estimates reported in Raikou and McGuire (2004) – which were 11901.01 (*se*=1061.36) and 10629.97 (*se*=510.00) for the conventional and intensive arms

[5] Estimated regression coefficients for the various models available on request.
[6] Estimated regression coefficients available on request.

respectively – reveals a close similarity which confirms that ordinary least squares regression results in biased estimates when the outcome variable is censored.

With respect to Lin's (2000) parametric approach that uses information only on the complete total costs, the resultant difference in mean cost between the trial arms is of the wrong expected direction. In addition, the estimated mean cost for the intensive group is much higher than expected. This pattern alters when the regression uses information on multiple cost observations on each patient obtained at a number of points in time over the study period. The latter approach results in estimates that are very close to its non-parametric counterparts, derived from the first Lin non-parametric method using information on individual cost histories, and even closer to the Bang and Tsiatis partitioned estimator, which again uses information on individual cost histories. Thus, the regression model which uses cost history information from all individuals results in a significant improvement compared to the parametric model which discards cost information from the censored cases. This was anticipated and confirms Lin's argument that the multiple time intervals approach improves efficiency by using information that is ignored by the complete costs approach. The estimates of the regression parameters for both Lin regression models and the naïve ordinary least squares regression are reported in Table 4.

Table 4. Estimated regression parameters for the naïve OLS and the Lin regression models

Estimator	**Conventional** Regression coefficients	Standard error	**Intensive** Regression coefficients	Standard error
Naïve OLS				
Const	-23980.37	13813.83	5647.46	8007.90
Age	262.47	161.68	57.55	100.61
Bmi	454.01	258.15	34.84	108.39
fpg	537.43	654.39	-176.55	231.98
Race	1783.51	1990.84	146.71	998.54
Sex	1545.94	2627.59	1802.75	1318.99
Lin complete costs				
Const	-21043.55	25522.02	32901.42	24882.84
Age	141.61	337.58	-211.60	315.09
Bmi	596.27	610.46	208.61	267.38
fpg	1099.66	1012.11	-979.99	753.35
Race	1424.06	4309.11	1740.02	2739.86
Sex	-206.40	8907.59	-2619.76	4142.53
Lin multiple intervals				
Const	-217.49	8723.92	12170.94*	4883.95
Age	-16.99	141.88	1.86	41.83
Bmi	127.08	211.20	23.84	108.31
fpg	1493.37*	634.84	148.15	187.03
Race	-247.16	1711.94	-863.83	630.64
Sex	139.01	3271.34	517.88	806.28

*significant at the 5% level

The coefficient estimates resulting from all these regressions indicate that the covariates have low explanatory power, although it should be emphasised that high standard errors on individual coefficients cannot be taken to mean that any particular model in its entirety has low *predictive* power. With respect to the individual coefficients in the Lin regression models, all are insignificant in the complete costs approach and significant only for fasting plasma glucose in the conventional group in the multiple time intervals approach. This finding is not surprising given that the set of available covariates has been determined by their clinical rather than economic importance. Of greater importance is the finding that the mean cost estimates derived from the multiple time intervals regression model appear much improved, being very close to the comparative non-

parametric estimates, relative to the respective cost estimates resulting from the complete costs regression. As both regressions use the same inverse probability weight in an attempt to account for censoring, the most likely explanation for this result is the increased cost information used in conjunction with the particular weight by the multiple intervals regression. In an attempt to explore the sensitivity of the mean cost estimates to the set of covariates included in the models, both Lin estimators were applied to the data incorporating fasting plasma glucose as the only covariate. The results together with the naïve ordinary least squares estimates are shown in Table 5 for the regression parameters and in Table 6 for the mean costs.

Table 5. Estimated regression parameters using fasting plasma glucose as the only covariate

	Conventional		Intensive	
Estimator	Regression coefficients	Standard error	Regression coefficients	Standard error
Naïve OLS				
Const	7367.78	5391.06	12142.42*	2171.18
fpg	602.40	624.86	-144.15	232.21
Lin complete costs				
Const	-466.98	9367.33	24497.55*	8912.96
fpg	1690.44	968.63	-891.05	996.83
Lin multiple intervals				
Const	4263.38	3963.33	13220.31*	1948.28
fpg	1270.71*	575.85	99.13	203.60

*significant at the 5% level

Table 6. Estimated mean costs from regression models using fasting plasma glucose as the only covariate

	Conventional		Intensive	
Estimator	Mean	Standard error	Mean	Standard error
Lin complete costs	13870.84	7060.85	16821.34	2419.14
Lin multiple intervals	15041.21	1578.42	14074.33	454.88
Naïve OLS	12477.19	1212.05	10900.58	560.30

The mean cost estimates resulting from the regressions in which fasting plasma glucose was the sole covariate are very similar to their respective counterparts derived from the models using the complete set of covariates. In this particular application, therefore, the choice of the set of covariates does not appear to have an impact on the resultant mean cost estimates. The inverse of the probability of an individual not being censored entering the estimating equations seems to be primarily responsible for the resultant predicted mean estimates. However, this particular weight alone is incapable of adequately adjusting for the loss of information when the level of censoring is too high as indicated by the poor performance of the complete costs regression. As was the case in the non-parametric analysis (Raikou and McGuire, 2004), the amount of available information on the cost history process proves as important as the probability weight, which adjusts the estimates for the information loss due to censoring.

4. Concluding comments

Parametric approaches provide a necessary alternative in deriving estimates of cost statistics in a number of circumstances, such as when interest lies in the assessment of individual covariate effects on cost or in the extrapolation of estimates beyond the observed study duration or to different patient populations. Inherent in all parametric approaches is the specification of a functional form for the relationship between the outcome variable and a set of independent variables. Naturally, a first candidate in this category would be the classical linear regression model with cost forming the dependent variable, but such an approach is known to yield biased

estimates when the outcome variable is drawn from a censored distribution. The naïve solution of estimating the regression parameters by completely discarding the censored cases from the estimation process is also biased, with the degree of bias increasing as the proportion of censored observations increases. This, together with the failure of parametric and semiparametric regression models traditionally used in the analysis of time to event data to account for censoring in the cost estimates due to dependent censoring between cost at event and cost at censoring, has led to a set of alternative regression methodologies within the context of parametric censored cost analysis.

The first of these methodologies, introduced by Carides et al (2000), adopts a regression approach where cost is modelled as a function of failure time and adjustment for censoring in the cost estimates is achieved through adjusting failure time for censoring. Consistency of the estimator is ensured if the regression model specifying the relationship between cost and failure time is consistently estimated. The second alternative, introduced by Lin (2000), uses a linear regression approach where cost is related to a set of individual covariates and adjustment for censoring in the cost estimates is performed through use of the inverse of the probability of an individual not being censored in the estimating equations. The method has been shown theoretically to derive consistent estimates of the regression parameters while accounting for the presence of censoring and is not restricted by the censoring pattern. Two estimators result from this approach. The first uses the individual total accumulated costs at the individual's point of failure or end of study, while the second makes use of multiple cost observations obtained on each subject at various points in time over the study period. The main advantage of the latter estimator is an increase in efficiency by allowing use of cost information that is not used by the preceding estimator.

Although the estimators' statistical properties have been studied theoretically, their performance under conditions of heavy censoring has not been assessed empirically. This issue is addressed in this study through applying the proposed estimators to a clinical dataset that exhibits high levels of censoring and comparing the resultant estimates with the respective estimates derived from the best non-parametric estimators previously applied to the same data (Raikou and McGuire, 2004). The main findings are as follows. The Carides et al estimator resulted in biased estimates for all parameterisations considered for the relationship between cost and failure time. The results suggested that the major source of bias was the high degree of censoring rather than a potential misspecification of the regression model as similar results were obtained under a number of alternative parameterisations for the relationship between cost and failure time. Given that under this approach bias in the cost estimates arises from bias in the estimates of the regression parameters, it is not surprising that the estimated coefficients do not reflect the true parameter values when their derivation was based on only 18% of the observed data which constituted the uncensored subset. Therefore, although such an approach is appealing on the basis that it attempts to model the time pattern of cost, it is of limited value at high levels of censoring. Given the potential value of methods that allow extrapolation of cost beyond the study period, development of parametric models that successfully do so under conditions of heavy censoring appears to be a fruitful area for future research. Concentrating on the Lin regression methodology, the approach using cost information solely from the complete cases yielded biased estimates of cost as expected given the limited amount of cost information entering the estimation process, while the approach using information on individual cost histories resulted in estimates that were very close to the ones derived from the best performing non-parametric methods, which also use information on the individual cost history process.

These findings provide further insight into the general issue of cost estimation in the presence of censoring in the following manner. Aside from identifying a regression methodology which performs well under extreme censoring conditions, the results of this analysis strengthen the validity of the main conclusion reached in the corresponding non-parametric analysis reported previously. The general finding emerging from this analysis is that incorporation of the inverse of the probability of an individual not being censored in the estimating equations is instrumental in deriving unbiased estimates of medical care costs under conditions of censoring. Nevertheless, the success of the approach is dependent on the amount of available information on the cost history

process, as this will in turn determine the degree of retrieval of cost information missing due to censoring. In circumstances where the level of censoring is high, knowledge of the history of the process under study proves a determining factor in the performance of the estimator. The implication for the design of a clinical study where data on medical resource use are collected is that effort should be made to record information on cost generating events on each individual at intermediate points in time over the study duration. The findings derived from the preceding analysis provide conclusive evidence in support of this requirement, with the value of the available information on the cost history process increasing as the degree of censoring increases.

5. References

Bang, H., Tsiatis, A.A., 2000. Estimating medical costs with censored data. *Biometrika* 87, 329-343.

Brouwer, W., Rutten, F., and Koopmanschap, M., 2001. Costing in economic evaluation. In *Economic Evaluation in Health Care: Merging theory with practice*, Eds. Drummond, M., McGuire, A. Oxford University Press.

Carides, G.W., Heyse, J., F., Iglewicz, B., 2000. A regression-based method for estimating mean treatment cost in the presence of right-censoring. *Biostatistics* 1, 299-313.

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

Dolan, P., 2000. The measurement of health related quality of life. In *Handbook of Health Economics*, Eds. Culyer, A.J., Newhouse, J.P. North-Holland, Amsterdam.

Dolan, P., 2001. Output measures and valuation in health. In *Economic Evaluation in Health Care: Merging theory with practice*, Eds. Drummond, M., McGuire, A. Oxford University Press.

Dranove, D., 1996. Measuring costs. In *Valuing Health Care*, Ed. Sloan, F.A., 61-75. Cambridge University Press.

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

Dudley, R.A., Harrell, F.E., Smith, L.R., Mark, D.B., Califf, R.M., Pryor, D.B., Glower, D., Lipscomb, J., Hlatky, M., 1993. 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, 261-271.

Etzioni, R., Feuer, E., Sullivan, S., Lin, D., Hu, C., Ramsey, S.D., 1999. On the use of survival analysis techniques to estimate medical care costs, *Journal of Health Economics*, 18, 365-380.

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.

Garber, A., 2000. Advances in cost-effectiveness analysis. In *Handbook of Health Economics*, Eds. Culyer, A.J., Newhouse, J.P. North-Holland, Amsterdam.

Green, W.H., 1997. *Econometric Analysis*. Prentice-Hall International.

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

Koul, H., Susarla, V., van Ryzin, J., 1981. Regression analysis with randomly right-censored data. *The Annals of Statistics* 9, 1276-1288.

Lancaster, T., 1990. The Econometric Analysis of Transition Data. *Econometric Society Monographs* 17. Cambridge University Press.

Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalised linear models. *Biometrika* 73, 13-22.

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

Lin, D.Y., 2000. Linear regression analysis of censored medical costs. *Biostatistics* 1, 35-47.

Raikou M., McGuire A., 2004. Estimating medical care costs under conditions of censoring. *Journal of Health Economics* 23, 443-470.

Sculpher, M., 2001. The role and estimation of productivity costs in economic evaluation. In *Economic Evaluation in Health Care: Merging theory with practice*, Eds. Drummond, M., McGuire, A. Oxford University Press.

Sen, A., 1982. *Choice, Welfare and Measurement*. Harvard University Press, Cambridge, Mass.

Tsuchiya, A., Williams, A., 2001. Welfare economics and economic evaluation. In *Economic Evaluation in Health Care: Merging theory with practice*, Eds. Drummond, M., McGuire, A. Oxford University Press.

Weinstein, M.C., Zeckhauser, R., 1973. Critical ratios and efficient allocation. *Journal of Public Economics* 2, 147-157.

Weinstein, M. C., Stason, W.B., 1976. *Hypertension: A Policy Perspective*. Harvard University Press, Cambridge.

Williams, A., Cookson, R., 2000. Equity in health care. In *Handbook of Health Economics*, Eds. Culyer, A.J., Newhouse, J.P. North-Holland, Amsterdam.

For further information on this or any of the Health publications contact:

Champa Heidbrink
Managing Editor
LSE Health
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Tel: + 44 (0)20 7955 6840
Fax: + 44 (0)20 7955 6090
Email: c.heidbrink@lse.ac.uk
Website: www.lse.ac.uk/collections/LSEHealth/