A comparison of methods to adjust for censored cost data under differing censoring mechanisms.

Tracey Young and Martin Buxton

Health Economics Research Group, Brunel University, Uxbridge Middlesex UB8 3PH

tracey.young@brunel.ac.uk

Presented at the Health Economics Study Group Meeting, Canterbury

9th to 11th July 2003

Introduction: It is widely accepted that ignoring the issue of censoring when estimating mean total costs results in biased cost estimates. Numerous techniques currently exist to adjust for censored cost data in order to derive a more accurate estimate of mean costs. The aim of this study was to compare existing methods for adjusting for censored cost data under different censoring mechanisms.

Methods: 726 "complete" cases from the UK cost-effectiveness of liver transplantation study were used to form a "complete" data set where the mean total study cost was known. This cohort of patients were used to simulate artificial data sets to compare methods for adjusting for censored costs across different censoring mechanisms – random, end-of-study, informative and partial censoring. The accuracy of methods was measured by comparing the difference between estimates and the "true" cost for the cohort and using sampling standard errors.

Results: The mean cost for the full data set, without censoring was £36,045 (standard error: 1,517). Of the 12 methods applied, Lin's method, where patient cost histories where known, consistently performed well across different censoring mechanisms (Difference from "true" cost ranged from: -£1,532 to £305). When the percentage of censored cases was low (10%) seven other methods also produced a reasonable estimate of mean total costs.

Conclusions: Lin's method for estimating mean costs in the presence of censoring out-performed all other methods across all censoring mechanisms. However, given that some of the methods are computer intensive, when the percentage of censored data is small it might be reasonable to ignore censoring.

Introduction

It is now a well-recognised fact that censoring occurs in cost as well as survival data. In survival analysis there exist established techniques, such as the Kaplan-Meier method or Cox's proportional hazard regression models to adjust for censored data within the context of the analysis. However, these methods will give biased estimates if applied to cost data, as the assumption of independence between censoring and costs is broken [Halstrom & Sullivan, 1998; Lin *et al*. 1997]. Therefore alternative methods need to be used.

Over recent years several methods have been proposed in the literature for deriving the mean total study costs in the presence of censoring [Lin *et al*. 1997; Lin, 2000; Carides *et al*. 2000; Bang & Tsiatis, 2000]. Some methods make use of the detailed patient cost histories sometimes recorded throughout studies and the remainder make use of each patients total study costs. Several studies have shown that these methods will produce an unbiased estimate of the mean total study costs providing that censoring occurs in a random fashion.

Censoring can occur for many reasons, for example patients may drop out of a study by moving from the area, the study may end before all patients' experience the event of interest or patients may withdraw from a study because of treatment side effects. This latter cause of censoring is known as informative censoring and can cause biased estimates, as censoring is dependent on the time spent in the study [Collett, 1994].

The aim of this paper was to compare the performance of the different methods for estimating mean total costs in the presence of censoring across different censoring mechanisms and to see whether any particular method consistently out performed the other methods. A cohort of patients from the cost-effectiveness of liver transplantation (CELT) study is used throughout this paper to illustrate the methods.

Methods

The CELT data set

Between January 1996 and December 1997 a total of 755 patients with end-stage liver disease were assessed for their suitability as liver transplant candidates at one of six Department of Health liver transplant centres in England. This data has previously been described in detail elsewhere [Longworth *et al*, 2003]. Briefly, all adult patients assessed for isolated liver transplantation at the six centres participating in the study were followed from point of assessment through to listing for liver transplant to 24 months post liver transplant operation. Detailed information on resource use at the transplant centre was collected during the study, as was information on patient survival and

clinical information, which was used as a measure of liver disease severity. Utilities were measured using the EuroQol EQ-5D which was administered as part of a postal questionnaire at three monthly intervals from point of listing until transplant and at 3, 6, 12 and 24 months post transplant.

The cohort of patients analysed here consists of 726 patients who were not censored within the study period who were followed for 27 months from point of assessment (this period represented the mean time the cohort spent on the waiting list (3 months) plus the 24 month follow-up period post liver transplant). The other 29 patients were excluded because they did not have 27 months worth of follow up data. The mean total cost and standard error for the "complete" cohort were calculated and these were compared with the estimated mean total costs from artificial data sets with censored data under the four censoring mechanisms for each of the censored cost methods described below.

Censoring Mechanisms

There are many reasons why censoring might arise and these may be classified into different types of censoring mechanisms. In this study differences in mean total study costs were estimated under random censoring, end of study censoring, informative censoring and partial censoring.

1. Random censoring

Random censoring occurs when patients are lost to follow-up for reasons that are independent of the event of interest. In the CELT study the event was death. Patients may drop out of the study at any time during the study period for numerous reasons, such as moving out of the study area or dying from causes unrelated to the study treatment (an example of this is being run over by a bus in a study investigating diabetes mortality).

Artificial data sets containing randomly censored data were created from the CELT cohort using simulation methods. Random number generators were used to select which patients were to be censored and at what time point the censoring would occur. Each patients study costs were then censored at the randomly determined censoring times. A censoring level of 10% was chosen and 3,000 simulated data sets were created.

2. End-of-study censoring

In the majority of studies patient cohorts are not recruited into a study at the same time point but over a period of time. Additionally, studies tend to run for a pre-specified time period, for example two-years. So it is possible that a proportion of patients not recruited on the first day of a study may have incomplete follow-up data over a two-year period. As with random censoring, simulation methods were used to create 3,000 artificial data sets for analysis, in these sets the time period

where censoring could occur was restricted to the last six months of the study data. A censoring level of 10% was chosen.

3. Informative censoring

Informative censoring occurs when patients drop out of a study for reasons that are related to the event of interest. For example a patient may withdraw from a study because their condition improves and this improvement is related to the treatment they are receiving. It is not always possible to determine the reason a patient has dropped out of a study so informative censoring may not be detected.

The results from the CELT cohort quality of life study were used in order to create data sets that simulated informative censoring. Two data sets were created: the first assumed that patients with low utility scores using the EuroQol EQ-5D would drop out of the study due to ill health and the second that patients with high scores would drop out of the study due to an improvement in their condition. In the first data set, patients whose EQ-5D score dropped below the 10^{th} percentile of the distribution (utility scores ≤ 0) at any time point during the study were censored at the point their score fell below this level, this resulted in 22% of the data being censored. In the second data set, patients were censored at the time point that their EQ-5D response exceeded the 90^{th} percentile of the EQ-5D distribution (utility scores ≥ 0.89, 28% of patients censored). For each data set patient study costs were set to zero from the point of censoring.

4. Partial censoring

Partial censoring may arise when patient's survival length and outcome are known but costs and quality of life information are only available up to a specific time point. Partial censoring may also arise when costs are collected at specific time points, for example 6 monthly intervals. Thus, a patient dying in month 8 of a study would have a known 6 monthly cost but a censored cost at 12 months.

A cohort with partially censored data was created by assuming that resource use and costs were collected at one time point only during the CELT study period, on 30 March 1998, 27 months after the CELT study first began. Any costs incurred after this time were set to zero and costs were censored as of this date (78% of the data was censored).

Estimating mean total costs in the presence of censoring

The majority of methods published in the literature, for estimating mean total study costs in the presence of censoring, are based on the Kaplan-Meier survival method [Kaplan & Meier, 1958]. Outlined below is a brief description of this methodology. Interested readers should refer to the original articles for further details of these methods.

1. Ignoring censoring, referred to here as the **available sample method,** will give an under estimate of total costs incurred beyond the point of censoring as these costs are ignored. However, estimating mean costs in the same way one would calculate mean age, for example, is still a frequently used method in cost-effectiveness studies.

2. Ignoring censored cases and estimating mean costs across the uncensored cases will also give a biased estimate towards the cost of patients with shorter survival times. This method is referred to as the **uncensored cases method**.

3. **The Kaplan-Meier (KM) cost method** is now recognised as giving biased estimates of mean total study costs [Lin, 1997; Halstrom & Sullivan, 1998]. Mean costs are estimated by calculating the area under the Kaplan-Meier survival curve, using a cost rather than time scale. Bias is caused when the assumption of independence between the censoring mechanism and costs is broken. Patients with lower costs will always be censored before patients with higher costs.

4. A natural extension to the Kaplan-Meier method is to apply **Cox's proportional hazard regression model** (the Cox cost method), on a cost scale, in order to adjust for factors that may effect patient costs. In addition to the assumptions of independent censoring there should be proportional hazards between different levels of fitted variables in the model, e.g. males and females, in order to obtain unbiased estimates.

   A Cox proportional hazard model was fitted to the "complete" CELT data set to determine what factors influenced patient costs. The proportionality assumption was checked for each variable. Age, transplant centre, liver disease group, gender and whether a patient received a liver transplant during the study were found to be significant predictors of total study costs. These variables were then fitted in models to estimate mean total costs for each of the simulated data sets for the four censoring mechanisms.

5. To overcome the problem of independence Lipscomb *et al.* proposed the **stratified Cox model** for estimating mean costs, where the study interval is divided into smaller time intervals (or strata) and a Cox proportional hazards model is fitted within each strata [Lipscomb *et al*. 1998]. The censoring mechanism differs slightly for this method as patient costs are censored within strata if they are incomplete for the full strata, reasons for censoring include death or incomplete cost collection due to patient censoring, patients with complete costs within each strata are not censored. The estimated costs are then summed across strata.

Strata of 6-month lengths were chosen for the CELT study, where the final stratum was of length 3 months. Variables for transplant centre, type of transplant (routine, emergency, re-transplant within 14 days of previous transplant, re-transplant greater than 14 days after previous transplant) and whether the patient received a transplant were fitted in the Cox models for all simulated data sets.

6. Lin *et al.* recognised the problems of biased estimates arising from using some of the methods described above and suggested two alternative ways of estimating mean costs in the presence of censoring [Lin *et al.* 1997]. The first method referred to here as **Lin's method, cost histories known (CHK)**, consists of dividing the study period into equally spaced smaller time intervals. In each interval the mean cost incurred, per interval, for all patients alive at the beginning of the interval are calculated. These mean costs are then weighted by the Kaplan-Meier survival probability of being alive at the start of the interval. The weighted costs are summed over time to obtain an estimate of the mean total study cost. For the CELT example, the study period was divided into 27 monthly intervals.

7. With **Lin's method where cost history are unknown (CHU)** it is unnecessary to have detailed information on patient cost histories, as only the patient's total study cost is used. As before, the study period is divided into equally spaced time intervals (1 monthly ones were chosen for the CELT example) and for each interval the mean total costs of patients dying during that interval is calculated and weighted by the Kaplan-Meier probability of dying during that interval. The costs for patients alive at the end of the study are included in the mean total costs for the final interval. The weighted values are then summed over the whole study period to obtain an estimate of mean total study costs in the presence of censoring.

8. Lin proposed a further two methods for estimating mean study costs in the presence of censoring based on ordinary least squares regression analysis [Lin, 2000]. The first of these methods is referred to here as **Lin's regression estimate CHU**. A weighted regression model is fitted to the total study costs of patients who died during the study or had complete costs to the end of the study period, and factors known to influence costs are adjusted for in the model. The weighting is calculated as the inverse of the Kaplan-Meier survival estimate, where the censoring indicator is reversed so that deaths are denoted as 0 (rather than 1) and censored cases as 1 and the Kaplan-Meier survival estimate is calculate in the usual way. The model is then applied to the entire study cohort, including the censored cases, to obtain an estimate of the mean total study costs.

9. Patient cost histories can also be used in this regression method by dividing the study period into smaller intervals and fitting a regression model to the study costs incurred in that interval.

The same covariates should be included in all models. The regression coefficients for each interval are then added together and the result applied to the full study cohort as before. The method is referred to as **Lin's regression estimate CHK**.

Transplant centre, disease group and whether the patient was transplanted or not were the variables included in the regression models for predicting mean total study costs using the total study costs for each patient for both of Lin's regression methods.

10. Bang and Tsiatis adapted methods for adjusting for censoring of quality adjusted life years and applied them to censored cost data [Zhao & Tsiatis, 1997; Bang & Tsiatis, 2000]. The **weighted method, CHU** estimates the mean total study cost from the cases that die during the study or those where complete study costs are available. As with Lin's regression method the Kaplan-Meier survival estimates are calculated using a reverse censoring indicator (death/complete study costs = 0, censoring = 1). For each patient who died or had complete study costs, their total study cost is divided by the individuals Kaplan-Meier survival probability and these are then summed over time and divided by the study sample size to obtain a mean total cost in the presence of censoring, the sample size should include the censored cases.

11. An extension to method 10, known here at the **weighted method, CHK** makes use of information on patient cost histories over the study period. As with previous methods the study time interval is divided into a number of smaller intervals, for the CELT example three monthly intervals were chosen. Within each period the interval costs of patients with complete costs for the interval or those who die during the interval are weighted by the Kaplan-Meier survival estimate, using reverse censoring. These estimates are summed over intervals and over time and then divided by the total sample size (726 for CELT) for the cohort to obtain an estimate of mean total costs.

12. The final method presented in this paper is **Carides' regression method.** It is assumed that a patients total study cost can be modelled from their survival times. An appropriate model, which should be fitted to patients with complete study costs or those who died during the study, is chosen to predict total costs from survival time, where the model chosen should be the one that best fits the data. The study period is then divided into smaller intervals, 1 monthly intervals were used in the CELT example. For each interval the mean expected cost, weighted by the Kaplan-Meier probability of dying, for those patient who die in that interval is calculated. For the final interval those with complete costs are included in the calculation. The mean values are then summed over the study period to obtain an estimate of the mean total study costs.

Results

Table 1 describes the demographic data for the "complete" set of 726 patients assessed during the study period for their suitability as liver transplant candidates. Of the total cohort, 529 (72.9%) patients' were placed on the waiting list for a liver transplant (the majority of patients were listed as elective (non-emergency) cases), 456 (62.0%) received a transplant and 363 (50.0%) were alive 27 months after first being assessed.

The mean total cost for the "complete" cohort was £36,045 and the standard error was 1,517. The median total cost was £27,166 with costs ranging from £393 to £311,873 across the sample. Costs in year 2 and 3 are discounted at 6% [NICE guidance, 2000].

Figures 1 to 5 show the estimated mean total study cost in the presence of censoring for the twelve different methods for estimating the mean study cost for random, end of study, informative censoring where patients with low EQ-5D scores censored, informative censoring where high scores are censored and partial censoring, respectively.

Lin's method where cost histories were used was, consistently, the most accurate estimate of the "true" cost of £36,045 across all five censoring mechanisms. No method gave consistently better mean cost estimates across all censoring mechanisms when information on cost histories was unknown. However, the method that gave the least accurate estimate of mean total cost was the Kaplan-Meier cost method where the estimate of mean total cost was over 3.5 times greater than the "true" cost across all 5 censoring mechanisms. A statistical comparison of the 12 different methods, after adjusting for censoring mechanism, showed that there was a significant difference in mean estimates across methods (Friedman's χ^2_{11} = 35.65, p < 0.001). If the methods based on survival techniques on the cost rather than survival scale are ignored (Cost Kaplan-Meier, Cost Cox, and Stratified Cox) as these methods all break the independence assumption then the difference in the mean cost estimates for the remaining 9 methods is no longer significant (Friedman's χ^2_8 = 14.40, p = 0.072).

For random censoring, end of study censoring and informative censoring where patients with high EQ-5D scores were censored all methods except the cost Kaplan-Meier, Cost Cox, stratified Cox and Lin's regression method where cost histories were assumed unknown predicted the mean total cost within £1,000. No method predicted the results within £1,000 when the censoring mechanism was informative and patients with low EQ-5D scores were censored. This is probably because the censored patients are likely to be the sicker patients in the cohort and thus incur the highest costs, as these patients are censored before the high costs are incurred most methods will thus underestimate the mean total cost. Only Lin's method with known cost histories predicted the mean

total cost within £1000 for the partial censoring mechanism. There was a statistically significant difference in mean total costs estimates across the five censoring mechanisms after adjusting for method (Friedman's χ^2_4 = 9.86, p = 0.043).

As expected the Cost Kaplan-Meier, Cost Cox and Stratified Cox methods gave poor predictions of the "true" standard error (1,517) where the Cost Kaplan-Meier and Cost Cox tending to overestimate it and the Stratified Cox method tending to underestimate it. All other methods gave an accurate measure of the "true" standard error term.

Table 2 presents results from the complete case method and the uncensored cases method across different levels of censoring. The estimates of mean total costs from the complete case method deteriorate as the proportion of censored cases increases, though the poorer estimates are still within £3,000 of the "true" cost. Estimates fluctuate slightly less using the uncensored case method, however, estimates of the standard error become less precise as the level of censoring increases.

Discussion and conclusions

This paper has shown that Lin *et al's* method for estimating total study costs in the presence of censoring when details of patient cost histories are known consistently estimates the mean total cost more accurately than the other methods illustrated here, across different censoring mechanisms. Lin's method with unknown cost histories, the full sample method and Carides' regression method tended to give better estimates of mean total costs if it was assumed that cost histories were not collected and only the total study cost per patient was available. Predictably, the three methods that broke the assumption of independence between costs and the censoring mechanism produced the poorest estimates of mean total costs.

The accuracy of the mean total study cost estimates will depend not only on the method chosen to estimate the mean costs but also on the proportion of censored cases. Although the mean total study costs did vary as the level of censoring increased for complete sample and uncensored case methods, the variation from the "true" mean estimate was not large. However, the authors need to extend the analysis to the other methods for estimating costs in the presence of censoring to see if this observation is generalisable. This analysis is currently being undertaken.

With partial censoring and informative censoring the percentage of censored cases (78%, 22% and 28%) depended on the time point at which the data was censored or the choices of level of EQ-5D score at which the data was censored. Though the estimates of mean costs were also very reasonable under these censoring mechanisms, varying these choices may produce differing

levels of differences between the methods. The authors are in the process of further work exploring how varying the proportion of censored cases varies the accuracy of mean total costs.

Several of the methods for predicting mean costs in the presence of censoring, Lin's non regression methods, Lin's regression method CHK, stratified Cox, weight method CHK, and Carides' method involved dividing the study period into a number of smaller intervals. The choice of intervals for these methods will result in differing degrees of accuracy and the optimum number and lengths of intervals will depend on the choice of method. For most of these methods preliminary work, varying the interval lengths, showed that choosing smaller interval lengths tended to give more accurate estimates. However, further work is needed to obtain information on the optimal choice of interval lengths for each method, and how this might vary across different data sets.

Some of the methods for estimating mean total costs could be considered to be fairly computer intensive, particularly those that involve manipulating patient cost histories into a number of time intervals. Manipulating data into intervals involves assigning each item of resource use to a cost interval, if the data set has collected resource use at a detailed level, as was the case for the CELT study, then this process can be very time consuming. Additionally, some methods require the use of the bootstrapping technique [Manly, 1997] for estimating the standard errors. For example, Lin's non-regression based methods and the weighted cost methods and this process requires additional computing time.

When patient cost histories are available, Lin's method will give the most accurate estimate of mean total cost in the presence of censoring, regardless of the underlying censoring mechanism. However, given that there was little difference in the mean cost estimates when patient cost histories were unknown it is worth considering using the naive methods of the complete sample estimate or ignoring the censored cases, which are less computer intensive methods, to obtain an estimate of mean total cost.

The four different censoring mechanisms all produce favourable estimates of mean total study costs of liver transplantation in the presence of censoring for all methods except the Kaplan-Meier cost method, the Cox cost method and the stratified Cox method. Even the estimates produced from informative censoring, where sicker patients, as measured using the EQ-5D, were censored, produced estimates within £4,000 of the "true" mean total study cost. In reality, censoring will not occur solely due to one type of censoring mechanism, as data may be censored due to random, end-of study, informative and partial causes all within one data set and it will not always be possible to determine what the underlying censoring mechanisms might be. The research needs extending to other data sets to see whether the results shown here using the CELT study can be

generalised to other data sets and to explore what might happen where the underlying censoring mechanism is unknown.

Table 1: Demographic patient characteristics

Age in years	Mean (SD)	49.5 (11.8)
	Median (IQR)	52 (43 to 58)
	Range	18 to 73
Gender (%)	Males	334 (46%)
Centre (%)	1	160 (22.0%)
	2	50 (6.9%)
	3	118 (16.3%)
	4	54 (7.4%)
	5	90 (12.4%)
	6	254 (35.0%)
Transplant group (%)	Not listed	197 (27.1%)
	Elective	396 (54.6%)
	Emergency	75 (10.3%)
	Re transplant < 14 days	7 (1.0%)
	Re transplant > 14 days	51 (7.0%)
Disease group (%)	Cholestatic	372 (51.2%)
	Parenchymal	204 (28.1%)
	Fulminant	78 (10.7%)
	Other	72 (9.9%)

Table 2: Mean total cost estimates (with standard errors) for the complete sample method and the uncensored cases method for random and end of study censoring, with censoring at 10%, 30% and 50%

		10% censoring	**30% censoring**	**50% censoring**
Actual cost		36,045 (1,517)	36,045 (1,517)	36,045 (1,517)
Random	Complete sample	35,359 (1,500)	34,263 (1,469)	33,072 (1,437)
	Uncensored cases	35,981 (1,671)	36,256 (1,892)	36,571 (2,336)
End of study	Complete sample	35,983 (1,514)	35,935 (1,511)	35,866 (1,507)
	Uncensored cases	36,794 (1,694)	36,718 (1,930)	35,037 (2,340)

Figure 1: Random censoring 10%

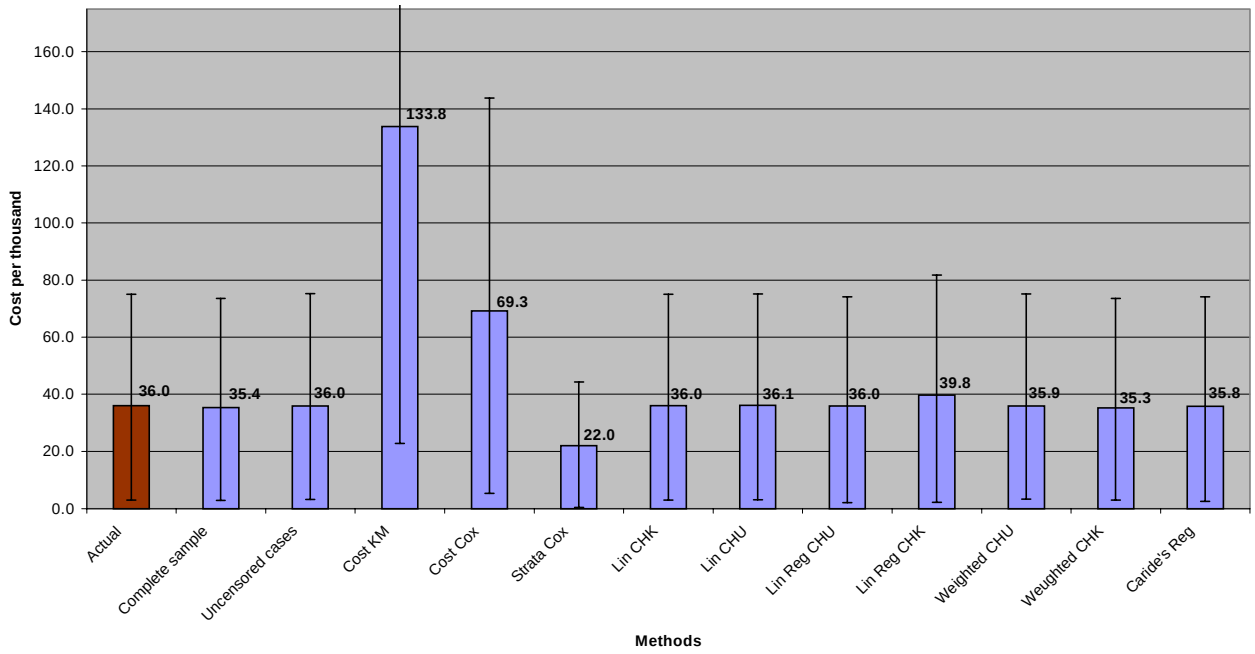


Figure 2: End of study censoring 10%

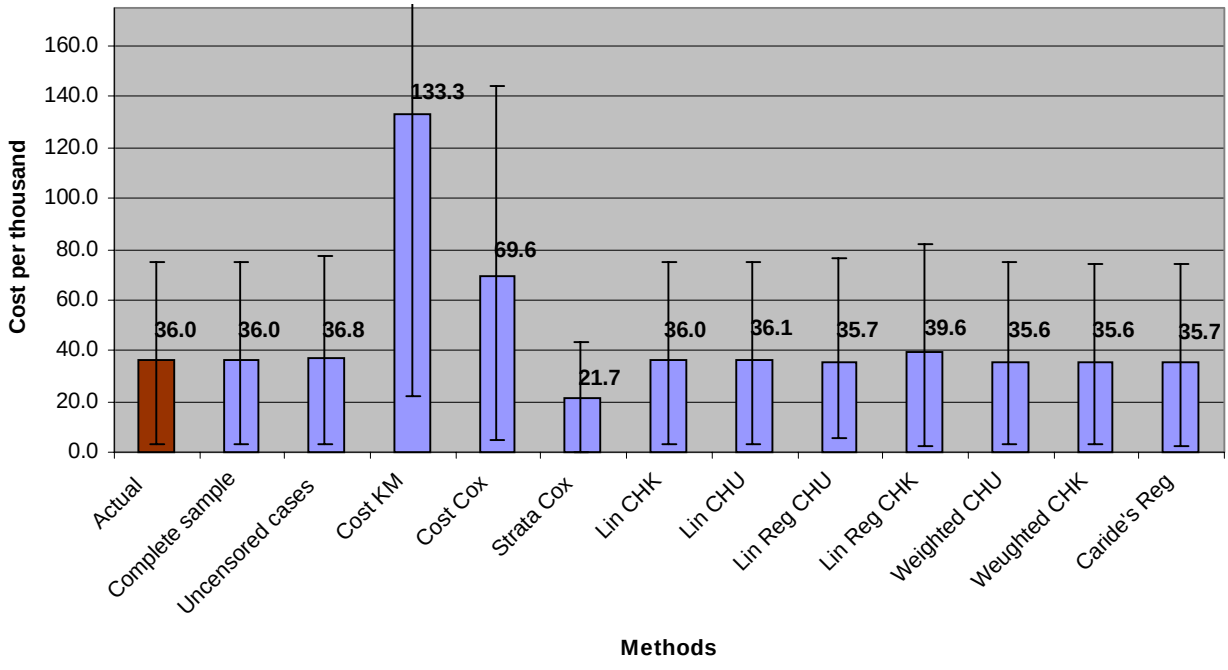


Figure 3: Informative censoring – Low EQ-5D scores censored

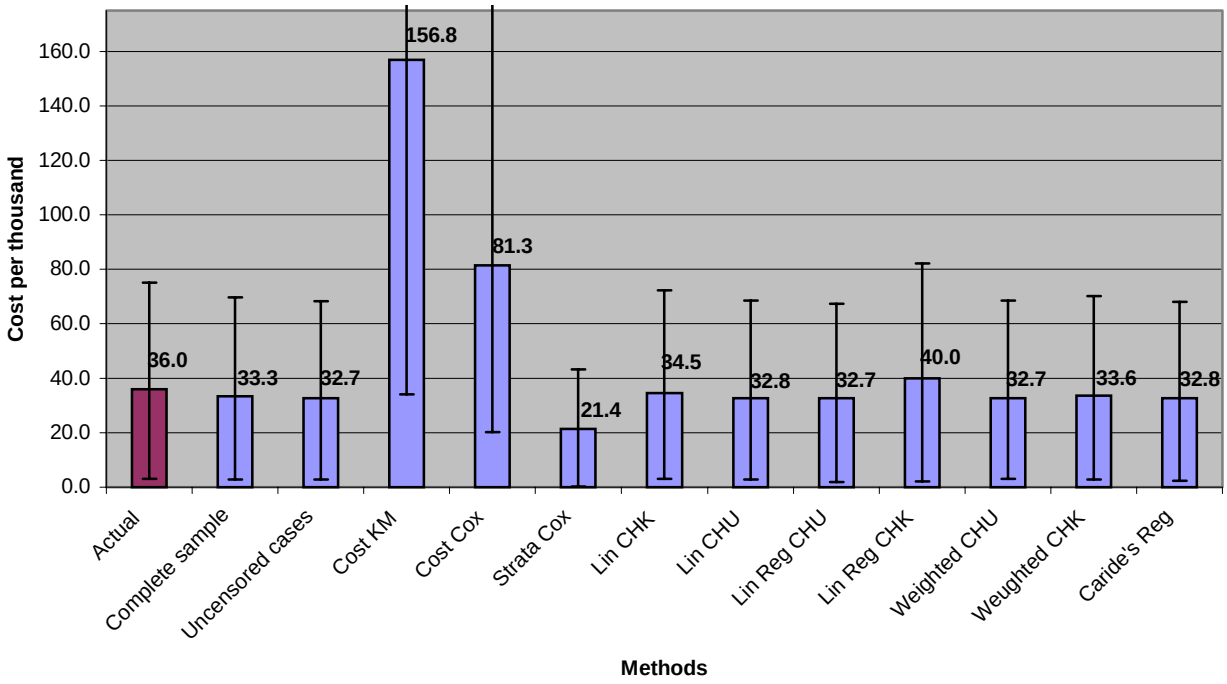


Figure 4: Informative censoring – High EQ-5D scores censored

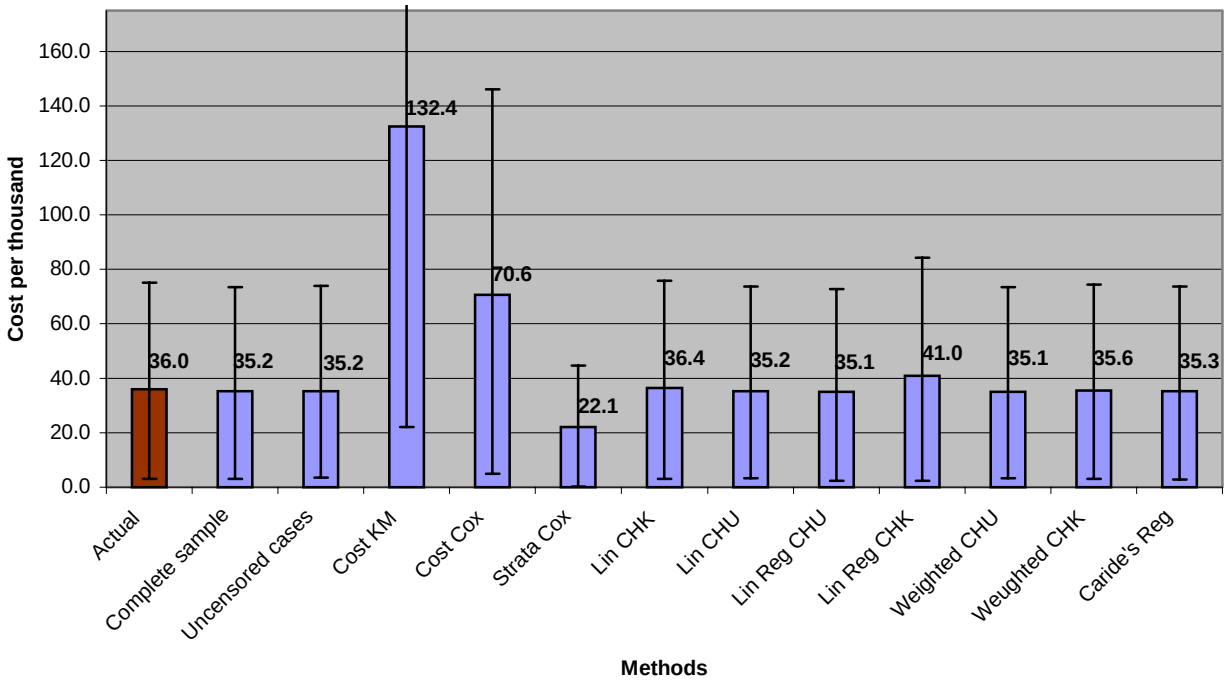
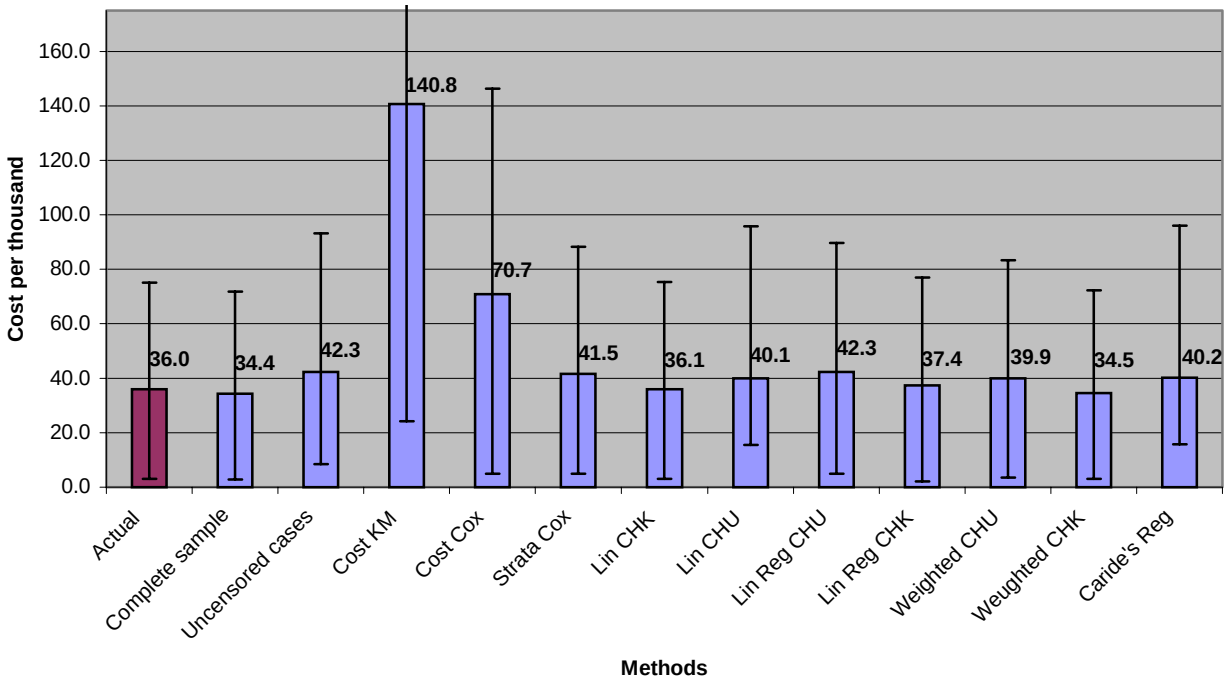


Figure 5: Partial censoring



References

Bang H, Tsiatis AA (2000) Estimating medical costs with censored data. Biometrika; 87(2): 329-343.

Cardies GW, Heyse JF, Iglewicz B (2000) A regression-based method for estimating mean treatment cost in the presence of right-censoring. Biostatistics; 1(3): 299-313.

Collett D (1994) *Modelling survival data in medical research.* Chapman & Hall, London.

Hallstrom AP, Sullivan SD (1998) On estimating costs for economic evaluation in failure time studies. Medical Care; 36(3): 433-436

Lin DY, Feuer EJ, Etzioni R, Wax Y (1997) Estimating medical costs from incomplete follow-up data. Biometrics; 53: 419-434.

Lin DY (2000) Linear regression analysis of censored medical costs. Biostatistics; 1 (1): 35-47.

Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar D (1998) Predicting the cost of illness: A comparison of alternative models applied to stroke. Medical Decision Making; 18 (Supplement): S39-S56.

Longworth L, Young T, Buxton M *et al.* (2003) Mid-term cost-effectiveness if the liver transplant programme of England and Wales for the disease groups. Submitted to Liver Transplantation.

Manly B (1997) *Randomisation, bootstrap and Monte Carlo methods in biology*. Chapman & Hall, London.

NHS National Institute for Clinical Excellence. Guidance for manufacturers and sponsors. April 2003. http://www.nice.org.uk

Zhao H, Tsiatis A (1997) A consistent estimator for the distribution of quality adjusted survival time. Biometrika; 84(2): 339-348.

Acknowledgements

We would like to acknowledge the assistance of the cost-effectiveness of liver transplantation (CELT) study team in the collection of the data on which this data emanates, in particular Louise Longworth. The CELT project has been funded by the England and Wales Department of Health Policy and Research Programme.

Any errors or omissions are the responsibility of the authors.