Estimating Mean Costs and Variances from Censored Trial Data

By

M. Raikou and A. McGuire

City University, London

Preliminary – do not quote without permission
Presented at the HESG, London, September 2001

Introduction

A particular problem in analysing clinical trial data relates to the handling of censored observations. While recognised methods exist to handle censored health outcome data there is less agreement over the handling of censored cost information. Recently Lin et al (1997) proposed two estimators of average treatment cost which allows for the presence of censoring. The authors show that their estimators minimise the bias induced by censoring if the study period is partitioned to narrow time intervals, and they are consistent if censoring occurs at the boundaries of the time intervals. They also show that the estimators are asymptotically normal and they derive their variance estimators drawing on the central limit theorem, Slutsky's theorem and the law of large numbers. A potential difficulty however associated with the practical application of the estimators is that the data may not justify the use of asymptotic theory to derive the variance. The purpose of this paper therefore is to derive the estimates of mean total cost and the associated variances using both approaches proposed by Lin et al, and subsequently compare the estimates of the obtained variances with the variances derived using the bootstrap method. Lin et al estimators were also derived for varying lengths of the time intervals of the partition, for varying durations of analysis, for varying levels of censoring and when high cost observations were excluded.

The paper continues with a brief description of the data used. This is followed with a methods section which outlines the rational and calculations employed by Lin et al in their estimation. This section is supplemented with a short discussion of the bootstrapping approach. A results section follows and the paper concludes with discussion.

Data

The data were taken from a randomised controlled clinical trial, the UK Prospective Diabetes Study (UKPDS) which followed-up 5,102 newly diagnosed type 2 diabetic patients with the aim of assessing the effectiveness of improved glucose control. Of these 5,102 individuals 3,867 entered the main randomisation and were allocated either to conventional policy (1,138) or intensive policy (2,729). The trial started in 1978 and ended in 1998 with a median follow-up period to death, the last date at which clinical status was known, or to the end of the trial of 10 years, ranging from 6 years to 20 years. For each individual in the study the trial collected information on both clinical effectiveness and resource use. The unit costs of hospitalisation and treatment medication were attached to the volume of resource use to calculate the total cost per patient per year directly from the trial data. These costs were then aggregated to give a total cost per patient for the whole trial period.

All estimators were applied to these trial data within each arm of the main randomisation, i.e. n=1,138 for the conventional policy arm and n=2,729 for the intensive policy arm.

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 the follow-up to death or the last contact date.

The failure event was all-cause mortality, resulting in 925 censored patients [81.3% censoring] (and 213 failures) in the conventional group and 2,240 censored patients [82% censoring] (and 489 failures) in the intensive group, by the end of the trial. Maximum time to death or last contact was 18.9 years for the conventional group and 19.5 years for the intensive group.

Table 0. Descriptive statistics of the data

	Conventional	Intensive
Sample size (n)	1138	2729
Censored	925 (81.3%)	2240 (82%)
Time duration of analysis (years): mean [range]	9.9 [0.01 to 18.934]	10.01 [0.05 to 19.463]
Time duration of analysis for failures (years): mean [range]	7.66 [0.12 to 16.73]	7.71 [0.05 to 18.41]
Time duration of analysis for censored (years): mean [range]	10.41 [0.01 to 18.934]	10.51 [0.197 to 19.463]
Total cost of failures: mean [range]	12,586 [220 to 145,549]	10,857 [20 to 110,921]
Total cost of censored: mean [range]	7,373 [119 to 189,242]	7,462 [149 to 97,121]

Methods

The main analysis estimates average treatment cost and the associated variance over the study period using both approaches proposed by Lin et al when time intervals are in years and in months to see whether the length of the intervals of the partition has an effect on the estimates. Bootstrap estimates of the standard error for both estimators as an alternative to the formulae for the variance were then calculated. As a secondary analysis the above methods are used with varying durations of analysis as the second method proposed by Lin et al appeared to result in relatively unstable estimates in the main analysis. The methods were subsequently applied excluding the three highest observed costs to see whether the estimates are sensitive to cost outliers. Finally one of the Lin et al approach was applied to an "artificial" dataset to assess its performance under controlled conditions and for varying levels of censoring.

Lin et al estimators

Notation and assumptions

Lin et al explore two approaches in estimating the mean total cost for a group of patients with the aim of minimising the bias induced by censoring. The first approach uses the sample mean of the observed costs from all study subjects and the second uses the sample mean of the observed costs from only the uncensored cases.

The assumption underlying both approaches is an extension of the one of *independent censoring* which, in the context of survival analysis, requires that at any follow-up time t patients cannot be censored because they are at unusually high (or low) risk of failure (dying). In the present setting, this definition is extended to require that at any follow-up time t patients are not censored because they will accrue unusually high (or low) costs.

The intuition behind the authors' idea is that bias will be minimised if censoring can be confined to the boundaries of the time intervals. To undertake this, Lin et al partition the study time period into narrow intervals to increase the probability that the censored data is pushed to the interval boundary. The approach can be outlined as follows.

The purpose is to estimate the mean total cost E = *E*(C), where *E* denotes expectation. Let C denote the total cost for a patient over the time period [0, τ), T be the survival time and U the censoring time. If T< τ, then C equals total cost up to T. The entire time period [0, τ) is divided into K intervals $[a_k, a_{k+1})$, (k=1,… ,K), where $a_1=0$ and $a_{K+1}=\tau$. The data typically consists of *n* independent observations of $(X, \delta, \tilde{C})$, where X = min(T, U), i.e. X is the last contact date, $\delta=I(T\leq U)$, where *I*(.) is the indicator function whose value is 1 if the argument is true (i.e. if the observation is uncensored) and 0 otherwise, and $\tilde{C}$ is the observed total cost, i.e. the cost accrued from the start of the follow-up to the last contact date X.

Two approaches to cost estimation
(1) uses the observed costs within the small intervals $[á_k, á_{k+1})$, (k=1….,K)
(2) requires only the observed total costs at the last contact dates

Using the cost histories (referred to as Lin 1 in the results tables)

The authors' first approach can be used to estimate E = *E*(C) when the cost histories are recorded in which case $\tilde{C}$ may be decomposed as $(\tilde{C}_1, \ldots, \tilde{C}_K)$, where $\tilde{C}_k$ is the observed cost observed over $[á_k, á_{k+1})$.

Assuming that the cost histories are recorded the mean total cost E is estimated by

$$\hat{E} = \sum_{k=1}^{K} \hat{S}_k \hat{E}_k \qquad (1)$$

where $\hat{S}_k = \Pr(T \geq a_k)$ is the probability of surviving to a_k consistently estimated by the Kaplan-Meier method as

$$\hat{S}_k = \prod_{j:t_j<a_k} \frac{n_j - d_j}{n_j}$$

and if $Y_{ki} = I(X_i \geq á_k)$ then $\hat{E}_k$ is the sample average of the observed costs over $[a_k, a_{k+1})$ among those who are under observation at the start of the interval, assuming that (extended) independent censoring as defined above holds, i.e.

$$\hat{E}_k = \frac{\sum_{i=1}^{n} Y_{ki} \tilde{C}_{ki}}{\sum_{i=1}^{n} Y_{ki}} \quad , k=1,\ldots,K$$

$\hat{E}_k$ is an unbiased estimator of E_k if censoring occurs at the *end* of the interval. The authors also suggest an alternative way of estimating E_k based on the exclusion of those who are censored during $[a_k, a_{k+1})$ from the calculation of the sample average $\hat{E}_k$. The resulting estimator will be unbiased if all the patients who are under observation at time a_k have the same probability of being censored during $[a_k, a_{k+1})$. This condition –which guarantees that the uncensored C_{ki}s are representative of all the C_{ki}s in the *k*th interval (the subscript *i* denotes individual *i*)- essentially requires that censoring occurs only at the *start* of the interval.

Clearly the bias diminishes as the intervals shrink and as the authors note both estimators of E_k are nearly consistent for narrow time intervals regardless of the censoring pattern.

The authors also show that for large *n* the estimator $\hat{E}_k$ is approximately normal with variance estimator given as

$$\hat{V} = \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{l=1}^{K} W_{ki}W_{li} \qquad (1a)$$

where

$$W_{ki} = \frac{\hat{S}_k Y_{ki}(\tilde{C}_{ki} - \hat{E}_k)}{\sum_{j=1}^{n} Y_{kj}} - \hat{S}_k \hat{E}_k \left\{ \frac{I(X_i \le a_k)\boldsymbol{d}_i}{R_i} - \sum_{j:X_j \le \min(a_k, X_i)} \frac{\boldsymbol{d}_j}{R_j^2} \right\},$$

and $R_i = \sum_{l=1}^{n} I(X_l \ge X_i)$

Not using the cost histories (referred to as Lin 2 in the results tables)

If the cost histories are not recorded the mean total cost E is estimated by

$$\hat{E}_T = \sum_{k=1}^{K+1} \hat{A}_k(\hat{S}_k - \hat{S}_{k+1}) \qquad (2)$$

where the survival probabilities S_k can be consistently estimated by the Kaplan-Meier method, with $\hat{S}_k - \hat{S}_{k+1}$ being the estimated Kaplan-Meier probability of death over the interval $[a_k, a_{k+1})$,

and $\hat{A}_k = \frac{\sum_{i=1}^{n} Y_{ki} C_i}{\sum_{i=1}^{n} Y_{ki}}$ where $Y_{ki} = I(á_k \le X_i < á_{k+1}, \boldsymbol{d}_i = 1)$.

i.e. $\hat{A}_k$ is the estimate of the mean total cost of those who are observed to die in $[a_k, a_{k+1})$ -again assuming extended independent censoring. This estimator may also be expressed as

$$\hat{A}_{K+1} = \frac{\sum_{i=1}^{n} Y_{K+1,i} C_i}{\sum_{i=1}^{n} Y_{K+1,i}} \text{ where } Y_{K+1,i} = I(X_i \ge \boldsymbol{t}).$$

$\hat{E}_T$ is an unbiased estimator of E if censoring occurs at the *end* of the interval or at the *start* of the interval. The former case is self-evident, the latter depends upon patients observed to die in the interval $[a_k, a_{k+1})$ being a random subset of all the deaths in $[a_k, a_{k+1})$. This is the case provided that independent censoring is maintained.

If censoring occurs in the interior of the interval, then $\hat{A}_k$ tends to be driven by the costs of the patients who die early in the interval because, given the same censoring distribution, larger survival times are more likely to be censored. However, if the interval is narrow, the costs associated with the early deaths of the interval are (stochastically) similar to those of the late deaths so that the bias of $\hat{A}_k$ will be small.

It is clear from the expressions for $\hat{A}_k$ (k=1, …, K+1) that the observed costs of the patients who are censored before $\boldsymbol{t}$ are not involved in any calculations and therefore need not be recorded. The costs of the other patients (i.e. those who are observed to die or whose censoring times equal $\boldsymbol{t}$) are allowed to be missing, but only in a completely random fashion. Y_{ki}=0 if C_i is missing.

For large *n* the estimator $\hat{\mathrm{E}}_T$ is approximately normal with variance estimator given as

$$\hat{V}_T = \sum_{i=1}^{n}\sum_{k=1}^{K+1}\sum_{l=1}^{K+1} W_{ki}W_{li} \qquad (2a)$$

where

$$W_{ki} = \frac{(\hat{S}_k - \hat{S}_{k+1})Y_{ki}(C_i - \hat{A}_k)}{\sum_{j=1}^{n} Y_{kj}} + \hat{A}_k(\hat{S}_{k+1}D_{k+1,i} - \hat{S}_k D_{ki}),$$

$$D_{ki} = \frac{I(X_i \le a_k)\boldsymbol{d}_i}{R_i} - \sum_{j:X_j \le \min(a_k, X_i)} \frac{\boldsymbol{d}_j}{R_j^2},$$

and $R_i = \sum_{l=1}^{n} I(X_l \ge X_i)$

The bootstrap

The bootstrap is normally used in the following circumstances. First, where an analytic estimate of the standard error of an estimator is too difficult or impossible to estimate; second, where it is believed that asymptotic theory provides a poor guide to the precision of a particular estimator and consequently an alternative is desired that may provide a better finite sample approximation.

The bootstrap estimates were obtained by drawing random samples of size N from the observed distribution of the sample and calculated the statistic, i.e. the Lin 1 and Lin 2 estimates of average treatment cost, across a large number of replications. Thus under the assumption that the observed distribution is a good estimate of the underlying population distribution the bootstrap produced an estimate of the sampling distribution of the statistic from which the standard error of the statistic was then calculated. All sets of bootstrap estimates were obtained for 200 and 1,000 bootstrap replications which are deemed adequate for the calculation of standard errors.

Artificial data

To assess the performance of Lin et al estimators under controlled conditions and for varying levels of censoring, an "artificial" dataset was constructed as follows. A sample size of 1138 patients was chosen for this "artificial" dataset to equal the smaller sample of the real UKPDS data (since one of the concerns for the validity of the methods is related to the sample size). Survival times were generated from a uniform distribution on [0, 10] years. The average 10-year cost serves as the parameter of interest, with the total cost for individual *i* is

$$Mi = M_i(0) + b_i T_i^L + \sum_{j=1}^{10} \boldsymbol{t}_{ij}(\min[\{T_i^L - (j-1)\}^+, 1]) + d_i I(T_i \le 10)$$

where $M_i(0)$ is the initial diagnostic cost, b_i is the deterministic annual cost, $\boldsymbol{t}_{ij}$ is the random annual cost for the jth year, d_i is the terminal death cost and $\alpha^+=\max(0, \alpha)$. For the distribution of each cost element, $M_i(0)$, b_i, $\boldsymbol{t}_{ij}$, d_i are assumed uniformly distributed on [5000, 15000], [1000, 2600], [0, 400] and [10000, 30000] respectively.

Various levels of censoring were considered: C_i being uniformly distributed on [0, 20] years, i.e. 25% censoring, [0, 12.5] years, i.e. 41% censoring, [0, 10] years, i.e. 51% censoring, [0, 9.5] years, i.e. 55% censoring and [0, 9] years, i.e. 57.5% censoring.

The components were generated independently and the estimator using cost histories was calculated for all levels of censoring.

Results

Table 1 presents the resulting Lin 1 & 2 mean and variance estimators where two sets of results are shown for each randomisation group. The 1st set of results was obtained when the time intervals of the partition are in years, and the 2nd set was obtained when the time intervals are in months with the monthly costs calculated as the annual cost/12.

Table 1. Lin 1 and Lin 2 estimators

	Conventional		Intensive	
	Mean	Variance (se)	Mean	Variance (se)
Subintervals in years				
Lin 1	14006.2	805921.4 (897.73)	13172	115977.5 (340.55)
Lin 2	12428	405678.2 (636.93)	16910.64	1021857 (1010.87)
Subintervals in months				
Lin 1	13771.35	1051866 (1025.60)	13078.02	133915.8 (365.95)
Lin 2	12530.39	446502.9 (668.21)	16926.22	1025211 (1012.53)

Comments

1. Lin 1 & 2: Length of intervals of the partition does not appear to have an impact
2. Lin 2: intensive > conventional which is not what is expected, probably because this method relies on a "reasonable" number of deaths in each sub-interval of the partition. In the UKPDS data, this is not generally the case.

Table 2 reports the resulting Lin 1 and Lin 2 estimates for different durations of analysis. This was undertaken to address the problem experienced above: Lin 2 estimator may be affected by the small numbers in the later sub-intervals (particularly of deaths but also of individuals censored after ô). So the time duration of analysis is altered.

Table 2. Lin estimators for different durations of analysis, i.e. for different values of t

	Conventional		Intensive	
	mean	variance (se)	mean	variance (se)
All time, i.e. **t =18.9** for conventional and **t =19.5** for intensive				
Lin 1	14006.2	805919.15 (897.73)	13172	115974.3 (340.55)
Lin 2	12428	405678.2 (636.93)	16910.64	1021857 (1010.87)
t =18 years				
Lin 1	13564.66	637264.8 (798.29)	12752.36	115604.6 (340.01)
Lin 2	15409.63	8588583 (2930.63)	12597.07	1251443 (1118.68)
t =17 years				
Lin 1	12831.22	442406.3 (665.14)	12295.81	105344.8 (324.57)
Lin 2	14785.19	2762098 (1661.96)	14031.57	710260.5 (842.77)
t =16 years				
Lin 1	11884.79	341022.3 (583.97)	11434.42	60366.32 (245.7)
Lin 2	13683.87	1477411 (1215.49)	12206.42	340539.1 (583.56)
t =15 years				
Lin 1	11258.03	262696.5 (512.54)	10750.22	44138.15 (210.09)
Lin 2	12381.43	974672.1 (987.25)	11434.13	231112.3 (480.74)
t =12 years				
Lin 1	8869.467	128128.5 (357.95)	8642.844	26886.65 (163.97)
Lin 2	9230.879	209875.5 (458.12)	8858.892	48627.31 (220.52)
t =18.9 years for conventional (i.e. all time for conventional)and **t =19 years** for intensive				
Lin 1	14006.2	805919.15 (897.73)	13011.87	116030.4 (340.63)
Lin 2	12428	405678.2 (636.93)	14217.32	5231386 (2287.22)

Comments

1. For τ=all time, i.e. main analysis, Lin2(conventional)<Lin2(intensive), and Lin1(conventional)>lin2(conventional)
2. For τ=18 years, Lin1(intensive)>Lin2(intensive)
3. For all other durations of analysis, Lin1<Lin2 within the same group, Lin1(conventional)>Lin1(intensive), and Lin2(conventional)>Lin2(intensive).

This appears to show that Lin 1 is stable and that Lin 2 is sensitive to small numbers of deaths in the sub-intervals and of individuals censored after τ.

Sensitivity of Lin estimators to high cost outliers

Lin 1 & Lin 2 were applied to the data after excluding the following highest cost patients from each group

Total cost, censoring status and time in study of excluded observations

Conventional	Intensive
189241.8 (censored at 8.643 years)	110921 (died at 5.448 years)
145548.8 (died at 10.729 years)	99928.88 (died at 9.973 years)
142953.6 (died at 6.573 years)	97121.06 (censored at 9.946 years)

Table 3. Lin 1&2 excluding the 3 highest observed costs from each group

	Conventional		Intensive	
	Mean	standard error	Mean	standard error
Main analysis				
Lin 1	14006.2	897.73	13172	340.55
Lin 2	12428	636.93	16910.64	1010.87
Excluding 3 cost outliers				
Lin 1	13583.07	865.32	13058.34	334.46
Lin 2	12078.4	580.88	16821.76	1009.15

Comment
High cost outliers do not seem to have an impact

Table 4. Lin: Bootstrap results for estimating the standard error

*When the time of analysis was the complete follow-up period (**t**=18.9 for conventional and **t**=19.5 for intensive)*		
	Conventional	Intensive
Replications 200		
Lin 1	823.4994	333.5156
Lin 2	8085.001	3784.319
Replications 1000		
Lin 1	927.3038	343.5167
Lin 2	7392.851	3837.986
When the time of analysis was 17 years for both conventional and intensive		
	Conventional	Intensive
Replications 200		
Lin 1	628.704	307.9546
Lin 2	1541.102	724.1773
Replications 1000		
Lin 1	670.0437	322.9126
Lin 2	1789.984	763.5229

Comments

1. Bootstrap estimates confirm the values for the standard error obtained from the formula for Lin 1
2. Lin 2 standard error estimates are confirmed when duration of analysis is restricted to 17 years which is when the method (Lin 2) becomes stable

Table 5. Lin 1 estimates based on the "artificial dataset" (n=1138)

	Mean	Standard error
No censoring	41144.5	
Censoring 25%		
Lin 1	39545.6	311
Censoring 41%		
Lin 1	37367.4	355.8
Censoring 51%		
Lin 1	35456.3	354
Censoring 55%		
Lin 1	34280	296.1
Censoring 57.5%		
Lin 1	33686	271.9

Comment
Lin 1 results in "worse" estimates as the level of censoring increases.

Conclusions

Overall the Lin 1 method appears to give better estimates than the Lin 2 method. This was expected given that Lin 2 approach relies on a "reasonable" number of deaths within each interval of the partition and is sensitive to the number of individuals censored after τ. These numbers were indeed very small towards the last sub-intervals in the UKPDS and the results for varying durations of analysis confirm the sensitivity of the method to these numbers. The results became stable, i.e. the direction of the differences between the various estimates becomes as expected when the duration of analysis was restricted to less than 18 years, because this increases the number of deaths and also the number of individuals censored at τ.

The bootstrap estimates of variance appear to give estimates which are very close to the ones derived using the formulae. This is the case for both Lin 1 and Lin 2 approaches (for Lin 2 the results are confirmed when the method becomes stable as expected), which could be interpreted in two ways: Fist, the theoretically derived asymptotic properties of the estimators appear to be empirically confirmed, and secondly, the bootstrap method gives reasonable approximation of the variances.

In addition, high cost outliers do not seem to have an impact on the estimates derived from both Lin et al approaches, and finally, when the Lin 1 approach (on the basis that it performed better than Lin 2) was applied to the "artificial" data to assess its performance under varying levels of censoring, the estimator appeared to perform "reasonably" well with the estimates approaching the "true" value as censoring decreases.

Bibliography

Cox, D. and Oakes, D., 1984, Analysis of Survival Data, Monographs on Statistics and Applied Probability, no. 21, (Chapman and Hall, London)

Efron, B. and Tibshirani R., 1993, An Introduction to the Bootstrap, New York: Chapman & Hall.

Etzioni, R., Feuer, E., Sullivan, S. et al, 1999, "On the use of survival analysis techniques to estimate medical costs", *Journal of Health Economics*, 18, 365-380

Fleming T. and Harrington D., Counting Processes & Survival Analysis, New York: Wiley.

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

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

UK Prospective Diabetes Study Group, 1998, "UK prospective diabetes study 33: intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes", *Lancet*, 352, 837-853