I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0132

CONFIDENCE INTERVALS FOR THE EXPENSES LINKED TO HOSPITAL STAYS :
USEFULNESS OF BOOTSTRAPPING SURVIVAL CURVES

Cl. BEGUIN and L. SIMAR

http://www.stat.ucl.ac.be

Confidence interval for the expenses linked to hospital stays:

Usefulness of bootstrapping survival curves

Claire Beguin*, Léopold Simar**

* Centre d'Informatique Médicale, Cliniques Universitaires Saint Luc, Université Catholique de Louvain

** Institut de Statistique, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

Address for correspondence:

C.Beguin
Cliniques Universitaires Saint Luc - Université Catholique de Louvain
Centre d'Informatique Médicale
10 avenue Hippocrate
1200 Bruxelles, Belgium
phone 32-2-7644711
fax 32-2-7644717
cbeguin@infm.ucl.ac.be

Research support from ``Projet d'Actions de Recherche Concertées" (No. 98/03--217) from the Belgian Government is acknowledged

Summary

The financing of hospitals in function of the pathologies uses the mean of the total expenses or length of stays by DRG after exclusion of extreme stays with trimming rules like $Q_1 - k*(Q_3 - Q_1)$ or $Q_3 + k*(Q_3 - Q_1)$ where Q_1 is the first quartile, Q_3 is the third quartile and k usually takes the value 1.5. In order to validate these techniques, we analyse the behaviour of the estimator of the survival curve of the total expenses obtained in Kaplan-Meier or Cox models, by using bootstrap techniques. The bootstrap method used by Efron [1] in a similar context seems well adapted to this problem. The objective is to find confidence intervals for upper percentiles of the time-dependent variable of the survival function, in our case the total amount spent for a hospital stay, and to verify if the upper bound $Q_3 + k*(Q_3 - Q_1)$ lies in this interval. These confidence intervals will be particularly useful for analysing the sampling variability of these bounds. Due to this variability, it appears that these bounds have to be considered with great care.

Keywords

Bootstrap; Survival analysis; Confidence Interval; Expenses of hospital stay; DRG

Introduction

Epidemiologists frequently used survival analysis techniques in estimating treatment effectiveness in the presence of censoring. This methodological issue receives more attention from economists when analysing costs of medical treatment [2]. These techniques have also been used to estimate the probabilities and the durations of short stay hospital and nursing home use [3]. Some refinements are added by Etzioni [4].

Our interest in survival analysis is motivated by recent changes in hospital financing in Belgium. Since 1995, in order to penalise hospitals with excessive length of stay (LOS), the Belgian Ministry of Health estimates an average LOS by Diagnosis Related Group (DRG) [5]. The estimation of the mean is made after the exclusion of extreme stays with rules like $Q_1 - k*(Q_3 - Q_1)$ or $Q_3 + k*(Q_3 - Q_1)$ where Q_1 is the first quartile, Q_3 is the third quartile and k usually takes the value 1.5. After 5 years of utilisation of this financing modulation, the Ministry of Health intends to change it and considers a global budget per hospital in function of the pathology.

In this context, we focus our analysis on the distribution of the total expenses generated by short stay hospital with a special attention to the high values of the total expenses. The distribution is estimated by survival analysis (Kaplan-Meier and Cox models). The bootstrap replications allow to define the confidence interval for the survival curves and to measure the bias of the survival curves estimated on the bootstrap samples. If the bias is weak, the estimation of the survival curve on the bootstrap samples will be considered as good. So, it could be used to define confidence interval for the percentiles of the 'time-variable', in our case, the total expenses for a particular stay, for some particular values of the survival and to verify if the upper bound $Q_3 + k*(Q_3 - Q_1)$ lies in this interval. These confidence intervals will be particularly useful for analysing the sampling variability of these bounds.

This paper is organised as follows: in the next section, we briefly present the data used as illustration of our method, then we describe the estimation method used and its bootstrap analogous, finally we comment the results and we conclude.

Data

The objective is to work with data routinely collected in hospitals. At each patient discharge a basic medical data set is collected. This set contains information such as age, sex, admission in emergency, destination of the patient after the hospital stay, length of stay (LOS), length of stay in intensive care unit (ICU), diagnoses and procedures performed. Diagnoses and procedures are coded with the ICD-9-CM classification allowing to group hospital stays in Diagnosis Related Groups (DRGs) [5].

Furthermore hospitals have a database with the total amount corresponding to all the medical acts performed during the hospitalisation on the base of the reimbursement by the INAMI (Institut National Assurance Maladie Invalidité). This amount is added to the price of a day of hospitalisation multiplied by LOS to obtain the total receipt of a hospital for a given hospitalisation. From the point of view of the Ministry of Health, it means the total expenses of a hospital stay. As pointed above, we analyse below the higher tail behaviour of this quantity.

The database used in this study contains data from 34 hospitals collected during the second semester 1988. Three DRGs are chosen: DRG 110 'Major reconstructive vascular procedures, without pump, with complications', DRG 122 'Acute myocardial infarction, discharge alive, with complicated cardiovascular diagnosis' and DRG 124 'Circulatory disorders, except acute myocardial infarction, with cardiac catheterization and complex diagnosis'. DRG 110 contains 323 stays and 10.5% of them ends with the death of the patient. DRG 122 contains 1643 stays without death and 976 stays are classified in DRG 124 and 4.3% of them finish with the death of the patients.

Method

The survival techniques used are the Kaplan-Meier estimator and the Cox model where the 'time to failure' is replaced by the total expenses and where dead patients are considered as censored. As the total expenses are influenced by the characteristics of the patients, the use of the Cox model provides more refined analysis.

First a survival curve is estimated with the Kaplan-Meier estimator. We retain the value of the total expenses corresponding to the survival at 99%, 95%, 90%, 75%, 50%, 25%, 10%, 5% and 1%. In order to derive confidence intervals for the survival, we use the bootstrap technique proposed by Efron [1] in this context. A comprehensive presentation of the bootstrap method can be found e.g. in [6]. The new samples are generated by bootstraping (2500 replications) on the pairs (total expenses, death) where the pairs are drawn independently with replacement. Survival curves are estimated on these replications so that for each retained value of the total expenses, 2500 survival measures are available. The confidence interval of the survival is estimated by the percentile bootstrap method and the bias by the difference between the average of the bootstrap values and the observed value of the statistic.

Since the percentiles of the total expenses are the main object of interest in our analysis, the problem can also be taken from the opposite side. A survival curve is estimated with the Kaplan-Meier estimator. The value of the survival at 99%, 95%, 90%, 75%, 50%, 25%, 10%, 5% and 1% are fixed. The survival curve are estimated on 2500 bootstrap samples. For each fixed value of the survival, 2500 measures of the corresponding percentiles of the total expenses are available. Due to the discrete nature of the values of the total expenses in the sample, the survival curve is a step function. Thus, some caution is to be taken as, moreover, the distribution of the total expenses presents a long tail at the right side. There is little chance that the estimated value of the survival function takes exactly one of the fixed values (for instance 5%). The estimated value of the survival function will more often be defined only for values either just below, say S^-, or just above, say S^+. According to the chosen value, we end up with two different estimates of the percentile. This motivates an alternative estimate based on the idea that the total expenses are continuous by nature. The idea is to interpolate linearly between S^- and S^+. The procedure works as follows for getting the percentile q_p. Let q^- (and q^+) be the percentile corresponding to the observed value S^- (S^+, respectively), then we have:

$$q_p = q^- + \left(q^+ - q^-\right) * \left(\frac{p - S^-}{S^+ - S^-}\right).$$

The linear interpolation will offer a more stable estimation of the percentiles than the two first methods, in particular, in the bootstrap approach, because it allows to find percentiles for the total expenses on a continuous scale rather than only on observed values. Some theoretical methods are available to estimate confidence intervals for percentiles but they are quite difficult to implement in practice and, as pointed above, due to the discrete nature of the data, they provide non exact coverage probabilities. For this reason, we use the bootstrap method [1, 6, 7].

If we want to take into account explanatory variables like age, sex, urgent admissions, number of ill systems and length of stay in ICU, we perform the same estimations with the Cox model. In the next paragraph, we summarise the values of these variables in the three DRG analysed.

In DRG 110, 82.4% of the 323 hospital stays concerns men and 42.1% of admissions in emergency. The mean of the explanatory variables of the Cox model is 67.9 year for age (median 68), 1.8 for the number of ill system (median 2) and 1.6 for the number of days in ICU (median 0). For the total expenses, the average is 312452 BEF (median 257530 BEF). 1643 hospital stays are classified in DRG 122, among which 74.7% concerns men and 74.9% of urgent admissions. The mean of the explanatory variables is 61.6 year for age (median 62), 1.2 for the number of ill system (median 1) and 2.1 for the number of days in ICU (median 0). The average of the total expenses is 118050BEF (median 92921BEF). The DRG 124 contains 976 hospital stays, 73.2% of them concerns men and 41.6% admissions in emergency. The mean age of these patients is 59.1 (median 60). On an average they have 1.2 number of ill system (median 1) and stay 1.3 days in ICU (median 0). For the total expenses the average is 123891BEF (median 74509BEF).

The bootstrap samples are drawn by independent sampling with replacement of the septuples (total expenses, death, 5 explanatory variables). In this approach, in order to estimate the survival curve, the values of the explanatory variables must be specified. In our analysis, we will define for each DRG, two typical individuals: one corresponding to a patient characterised by a low value of his expenses (denoted by individual 1) and the other (denoted individual 2) characterised by high value of his expenses. Other types of individual could be also analysed (like a 'median' patient) but we preferred to choose these two 'types' for better analysing the influence of the explanatory variables.

In DRG 110, individual 1 is a man of 65 years, with 1 ill system, planned admission and no stay in ICU, and individual 2 is a man of 75 years, with 2 ill systems, urgent admissions and staying 4 days in ICU. In DRG 122, individual 1 is a man of 62 years with the same characteristics as individual 1 of DRG 110. Individual 2 is a woman of 62 years with 2 ill systems and urgent admissions but no stay in ICU. In DRG 124, a woman of 60 years, with 1 ill system, planned admission and no stay in ICU represents the individual 1 but individual 2 is a man of 60 years with 2 ill systems, urgent admissions and staying 2 days in ICU.

Results

Figure 1 shows the Cox survival estimation for the two individuals of DRG 124 as well as the Kaplan-Meier estimation. The survival curves obtained for the individuals 1 and 2 in the Cox model are quite different from the survival curve estimated by the Kaplan-Meier method. As expected, we see that the Kaplan-Meier survival curve is in between the two extreme survival curves obtained by the Cox model for two 'extreme' individuals: individual 1 with low expected cost on the left and individual 2 with high expected costs on the right. The same kind of figures are obtained for the two other DRGs but not provided here to save place.

The bias of the survival curve obtained on the 2500 bootstrap samples is estimated for each DRG. The bias estimate is the difference between the average of the bootstrap values and the observed value of the statistic [6]. Table I shows the bias of some percentiles of the survival curve estimated by a Cox model on 2500 bootstrap replications for individual 2 of DRG 110. These values were the highest values observed across all the analysis we have performed (whatever the DRG analysed and the model used, Kaplan-Meier or Cox). The highest bias value is observed for the median and is equal to 0.0073. This particular behaviour for the median is also observed by Efron [1]. Globally, the bias of the estimations could be considered as being weak. This was the case across all the analysis we have performed: so that the survival curve is rather well estimated either by the Kaplan-Meier or Cox model. Now we can use the survival curves to estimate the main object of our interest, i.e. the percentiles. As explained in the preceding section, we use the bootstrap technique to provide confidence intervals for these percentiles.

Tables II and III show the means and the confidence intervals at 5% of the total expenses for survival at 10%, 5% and 1% when the value of the survival fixed is just before or just after 10%, 5% and 1%, or estimated by linear interpolation. Table II concerns the total expenses of DRG 110 when survival is estimated by the Kaplan-Meier model. Table III deals with the total expenses of individual 2 of DRG 124 when the survival is estimated by the Cox model. As expected, estimating the survival by linear interpolation decreases the upper bound of the confidence interval and increases its lower bound. It is particularly true for the values of the total expenses corresponding to the survival at 5% and more important for the values corresponding to the survival at 1%. This is observed for each DRG analysed.

In order to appreciate the sampling variability of the estimated percentile of the total expenses, we display in figure 2 the length of the confidence interval as a function of the bootstrap means of its corresponding total expenses. This is done for the survival at 50%, 25%, 10%, 5% and 1% (using the linear interpolation). The lengths of these confidence intervals increase with the level of expenses. The longest interval are observed for the values of the total expenses of the survival at 5% and 1%. This behaviour was observed for all the analysis we performed whatever the DRG taken into account. This is not a surprise since it is more difficult to estimate tail behaviour of random variables, due to the lack of observations in the tails. As mentioned in the introduction, for computing an average cost of a stay, the Ministry of Health excludes the extreme stays with rules like $Q_1 - k*(Q_3 - Q_1)$ or $Q_3 + k*(Q_3 - Q_1)$. If k takes the values 1.5, 2, 2.5, or 3, the value of the upper bound is quite similar to the value of the total expenses corresponding to percentiles obtained by the Kaplan-Meier survival at

10% or less (for all DRG analysed). So this upper bound has a great sampling variability and it should be used with great care.

Conclusions

The financing of hospitals in function of the pathologies uses the mean of the total expenses after exclusion of extreme stays with rules like $Q_1 - k*(Q_3 - Q_1)$ or $Q_3 + k*(Q_3 - Q_1)$. The estimation of the bootstrap confidence interval of the percentiles for the total expenses obtained through the survival function allows the measure the variability of these percentiles. As the upper bound $Q_3 + k*(Q_3 - Q_1)$ corresponds to the percentiles of the total expenses having the largest variability, the techniques based on the interquartile range and similar techniques should be used with great care, in particular, by taking their variability into account.

References

1. Efron B. Censored data and the bootstrap. *JASA* 1981; **76**: 312-319.

2. Fenn P., McGuire A., Phillips V. et al . The analysis of censored treatment cost data in economic evaluation. *Med Care* 1995; **33**: 851-863.

3. Liang J., Liu X., Tu E., Whitelaw N. Probabilities and lifetime durations of short-stay hospital and nursing home use in the United States, 1985. *Med Care* 1996; **34**: 1018-1036.

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

5. Fetter R., Shin Y., Freeman J.L., Averill R.F., Thompson J.D. Case-mix definition by Diagnosis Related Groups. *Med Care* 1980;**18** (2): 53p.

6. Efron B., Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, Inc: New York, 1993

7. Tambour M. and Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. *Health Econ* 1998; **7**: 143-147.

Table I. Bootstrap estimation of the bias of the survival function for individual 2 in DRG 110

% stays failed	DRG 110 N = 323, 10.5% of censored observations		
	Total expenses *	Survival **	Bias
1	92499	0.9880	0.0043
5	130907	0.9491	0.0014
10	161572	0.8997	0.0026
25	217042	0.7474	0.0057
50	326653	0.4988	0.0073
75	497537	0.2453	0.0062
90	683210	0.0990	0.0029
95	864683	0.0478	0.0003
99	1540155	0.0031	0.0010

* values of the total expenses corresponding to the survival at 99%, 95%, 90%, 75%, 50%, 25%, 10%, 5% and 1% in the original sample.

** values of the survival estimated on the original sample.

Table II. Mean and confidence interval of total expenses after bootstrap replications for survival at 10%, 5% and 1% for DRG 110 (Kaplan-Meier model)

	Mean	Confidence Interval Lower bound	Upper bound
Survival at 10%			
q^-	572969	514648	667830
q^+	587758	516682	687402
q_p	580364	515665	675520
Survival at 5%			
q^-	728926	598166	858702
q^+	762040	660382	864683
q_p	745783	632998	861693
Survival at 1%			
q^-	1030820	799034	1540155
q^+	1259335	858702	1809143
q_p	1145078	855553	1674649

q^- = percentile corresponding to S^-, value just below p

q^+ = percentile corresponding to S^+, value just above p

q_p = percentile obtained by linear interpolation

p = 0.10, 0.05, 0.01

Table III. Mean and confidence interval of total expenses after bootstrap replications for survival at 10%, 5% and 1% for population 2 of DRG 124 (Cox model)

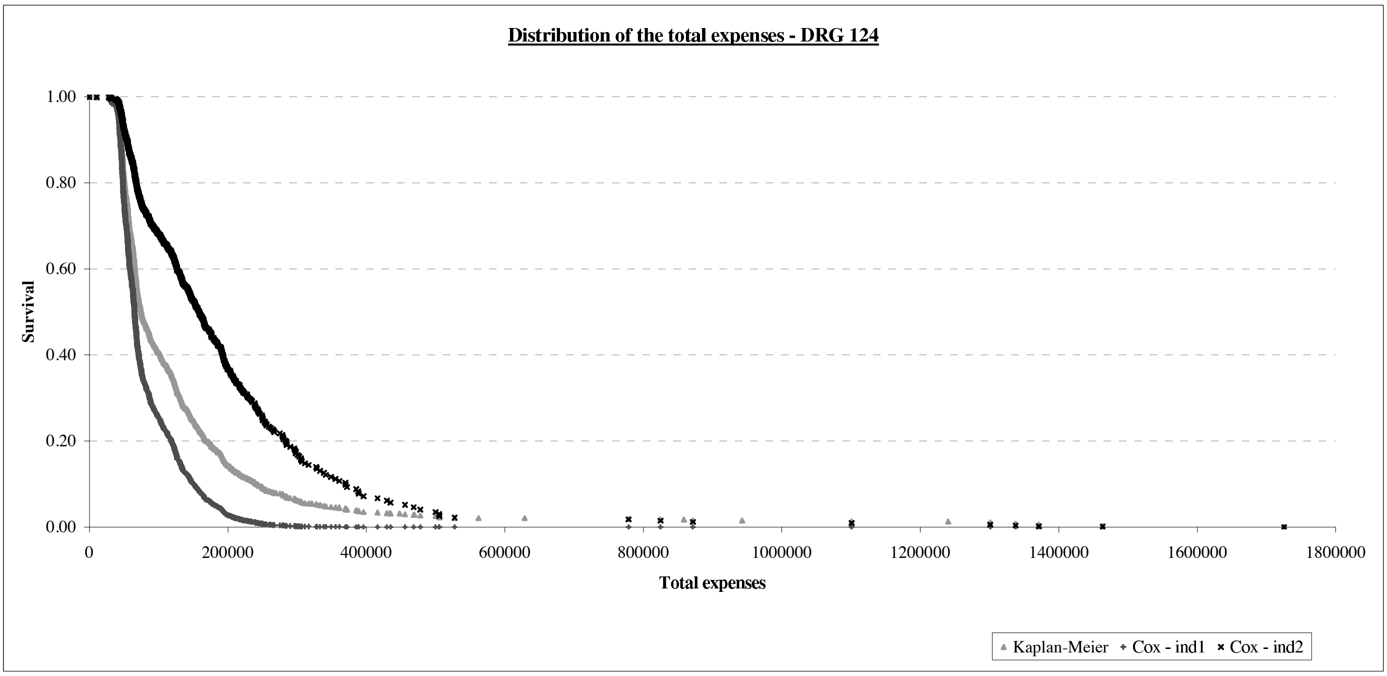
	Mean	Confidence interval Lower bound	Upper bound
Survival at 10%			
q^-	359420	306780	429720
q^+	368940	316543	434707
q_p	364180	313862	432214
Survival at 5%			
q^-	436504	369549	505409
q^+	463544	371917	778575
q_p	449924	370733	653022
Survival at 1%			
q^-	819801	455680	1463387
q^+	991335	477843	1725273
q_p	905568	466762	1594330

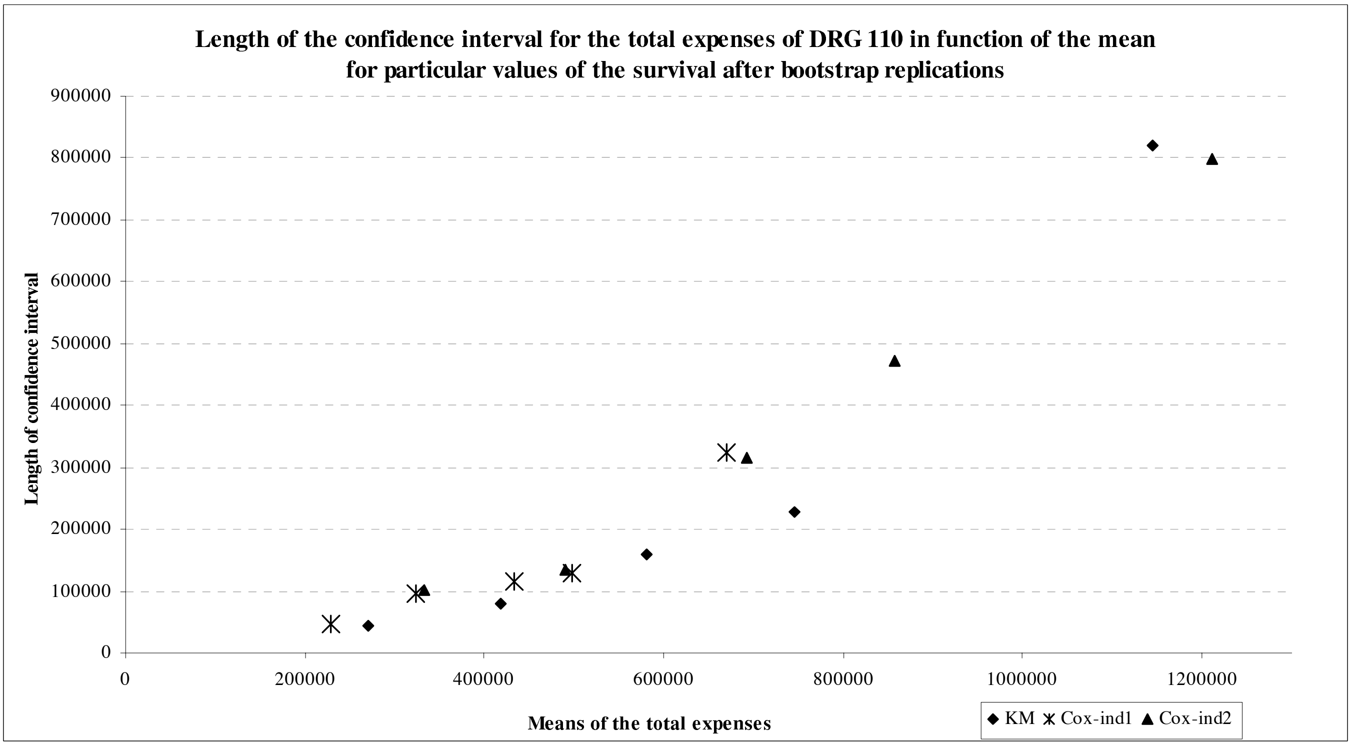
q^- = percentile corresponding to S^-, value just below p

q^+ = percentile corresponding to S^+, value just above p

q_p = percentile obtained by linear interpolation

p = 0.10, 0.05, 0.01


Distribution of the total expenses - DRG 124
1.00
0.80
0.60
0.40
0.20
0.00
Survival
0
200000
400000
600000
800000
1000000
1200000
1400000
1600000
1800000
Total expenses
Kaplan-Meier
Cox - ind1
Cox - ind2


Length of the confidence interval for the total expenses of DRG 110 in function of the mean for particular values of the survival after bootstrap replications
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
Length of confidence interval
0
200000
400000
600000
800000
1000000
1200000
Means of the total expenses
KM
Cox-ind1
Cox-ind2