Paper 284

ANALYZING ONCOLOGY PATIENT HEALTH CARE COSTS USING THE SAS SYSTEM

Jeff A. Sloan, Steven S. Cha, Judith L. Wagner, Steven R. Alberts, Jed Lindman, Cristine Allmer
Mayo Clinic, Rochester, MN 55905

ABSTRACT

The advent of managed health care has focused attention on the methodology for analyzing patient health care costs. Standard methods of analysis apply to a varying degree, but there are unique challenges inherent in constructing and analyzing cost data. Adjustments in basic statistical procedures are necessary to account for changes in the value of money over time, marked skewness due to catastrophic costs and censoring from both left and right. A number of authors have developed novel approaches for the analysis of patient cost data (Fenn et al, 1996; Siegel et al, 1996; Zhou et al, 1997; Etzioni et al, 1998). We present an array of alternative statistical machinations for cost data achieved through varied application of the SAS system. We use a recently completed study which compares the relative health care costs of patients who participated in oncology clinical trials to costs for patients who did not enter onto a clinical trial to demonstrate the various methods of analyzing cost data via a series of SAS macros.

INTRODUCTION

The number of clinical trials that include examination of health care costs has risen dramatically in recent years (Adams et al, 1992). Managed care has brought with it a number of research opportunities due to the need for accurate assessment of the relative efficiency of treatment alternatives. Clinical trial statisticians have been called upon to provide the analytical machinery to answer these important issues. In this paper, we explore the alternative analytical approaches in a series of SAS macros. We demonstrate the macros using a recently completed Mayo Clinic study comparing the health care costs for a sample of patients who participated in oncology clinical trials to a sample of patients who did not participate in any clinical trials. The data presented here are intended only as a vehicle for discussing the various alternative analytical procedures one can carry out on health care cost data. Figures and analytical results contained herein are for exemplary purposes only and should not be construed as representing the final study results. For a detailed discussion of the findings from the example study, the reader is referred to the primary publication by Wagner et al (1999).

EXAMPLE STUDY PRELIMINARIES

We will explore the various analytical procedures by demonstrating them on a study which was intended to explore the differences in health care costs experienced by oncology patients who participate in clinical trials versus patients who do not go onto any clinical trials. A cohort design was used for this study to compare all patients entering phase II/III oncology clinical trials at the Mayo Clinic in the period 1988-1994 with cancer patients who were treated at the Mayo Clinic during the same period but who did not participate in clinical research. The primary endpoint was the lifetime medical costs from all sources of care following date of entry onto cancer clinical trial to death or study termination (December 31, 1995). The analysis was made possible by a database created at Mayo which contains all medical billing data for medical care providers in Olmsted County, MN. A standardized unit cost was assigned to each detailed billing item, and the unit costs were adjusted to national norms in 1995 constant dollars.

Producing samples that can be compared with a reasonable degree of control over potentially confounding factors can be a challenge in retrospective studies. In our case, we followed a two-stage matching procedure. We first matched the roughly 1,200 patients who went on an oncology phase II or phase III clinical trial at Mayo over the specified period with a sample drawn from over 12,000 oncology patients matched for age, sex, disease site, disease stage and date of diagnosis. We subsequently produced more stringent matching criteria to include metastasis status, the ECOG performance score and a review of the patient's chart to establish that the patient would have been eligible to participate in the candidate clinical trials in which their counterpart matching case actually participated. The second stage matching produce 61 matched pairs for analysis. The advantages of the more precise matching have to be weighed against the reduced sample size obtained. The two-stage process allows for an assessment of the study's findings in the light of varying matching criteria. For the remainder of this paper we will present results for the matched pairs' data, although the methods were applied to the first stage matching samples as well.

ANALYTICAL APPROACHES: BASIC

There is, at present, no generally accepted statistical approach for cost data that is considered to be optimal (Siegel et al, 1996; Zhou et al, 1997; Grieve, 1998). As such, analysis of the cost data is best carried out in a number of complementary ways in order to perform concurrently a sensitivity analysis relative to the assumptions involved in each statistical procedure. Descriptive statistics can be used to assess the distributional assumptions and the need for transformations. With most cost-related data, the distributions are markedly skewed to the right due to a small number of observations with very high costs. The use of nonparametric procedures provides an alternative method of assessment to the traditional means-based procedures.

Logarithmic transformations have long been a staple of econometric analysis (Zhou et al, 1997). This approach makes the interpretation of the data slightly more involved, and the artificial induction of normality that is the goal of the approach is often not obtained, but this method does serve as a useful sensitivity analysis.

Discounting costs over time to their value in a reference period is another standard method applied to cost data (Siegel et al, 1996). This necessitates the creation of a further cost variable for analysis.

Although the primary endpoint of interest is the total health care cost incurred by each patient (discounted or undiscounted), there are numerous other cost variables that can be created to provide a more complete picture of the health care cost profile. Other cost variables can be created and analyzed by calculating the cost per month of follow-up, costs incurred in the first month, first three months, and first, second and third year of follow-up from the index date. The date of death can also be used as a time reference and costs can be calculated for costs incurred in the last month, three months, six months and last year of life.

ANALYTICAL APPROACHES: NOVEL

Novel analytical approaches that have been developed only recently specifically for cost data were also applied. Although the reliability of these new procedures has yet to be determined we felt it was important to include these approaches in our SAS code to provide a more complete analysis of the data. It should be noted that there remain contradictory recommendations in the literature for the appropriate analysis of cost data beyond relatively simplistic traditional methods. As such, the results from the new novel approaches need to be interpreted with caution.

One novel approach that was suggested by Fenn et al (1996) involved using survival analysis methods to compare cost profiles across treatment groups. While this method has some appealing graphical and intuitive characteristics, it has come under fire in the literature regarding the inherent bias of the Kaplan-Meier estimates involved (Hallstrom and Sullivan, 1998). An alternative approach (Lin et al, 1997), provides a method for adjusting the cost incurred by the Kaplan-Meier estimate of survival. The mathematical characteristics of this approach have yet to be fully explored. The basic idea is to take into account the fact that not all patients will provide complete cost data for all time periods due to a myriad of reasons. Simply accumulating costs for such patients, therefore, would underestimate the true total costs. The problem is analogous to the incomplete data problem in time-to-event studies, such as the classic example of survival analysis. In traditional survival analysis, such data are declared to be censored observations and summary estimates are adjusted accordingly.

The Kaplan-Meier Sampling Average (KMSA) approach to patient cost data is constructed by obtaining the average cost and Kaplan-Meier survival estimate for each time period (in our example, thirty day periods). The sum of the product of these two components becomes the KMSA estimator of total cost for the sample. In essence, the survival estimates are used as a weighting function for the cost data. To date there are no closed forms for the standard deviation, confidence intervals or hypothesis tests for the KMSA estimators. In his appendix, Lin et al (1997) offers an analytic formula for calculating the variance, but it involves arduous and, as of yet, untenable numerical programming. As such, we used straightforward application of the bootstrap method to obtain variance estimates for the KMSA cost variables. We applied the KMSA approach to the raw costs, logged costs and discounted costs.

RESULTS

A total of 61 matched pairs (cases and controls) formed the final dataset for analysis. Paired comparisons formed the primary basis of analysis involving the intra-case differences in costs. Summary statistics for total costs are given in Table 1. The first noticeable result is that the figures in the columns of discounted

costs are very close to those of the raw costs. Results based on the logarithmic data (not shown) provided similar results.

The 61 patients on clinical trials cost roughly $100,000 more than the 61 patients who did not participate in a clinical trial. Both the mean and median costs for patients on clinical trials were higher, although not statistically significant by either paired t-tests or Wilcoxon signed rank procedures. Results did not change when total costs were adjusted for the amount of time each patient was followed/lived. In terms of raw total costs over the study period from the index date, patients on clinical trials cost an average of $2,467 more than those who did not enter a clinical trial. The average total cost per patient was slightly over $40,000. The estimated increase in cost for a patient to go on a clinical trial is hence roughly 5.9% of the average cost for patients who do not go on a clinical trial. The cost per month of follow-up is roughly 10.8% more for the 61 patients who went on clinical trials. The median difference was in excess of $7,000. Variability among the cases was marked, however, with some cases costing over $200,000 more than their matched counterparts who did not go on study while other cases cost over $200,000 less than their non-trial matched patient (Figures 1 and 2).

For every month that a patient is alive and available to follow-up, the 61 patients on study cost a total of just over $15,000 more than their nonstudy counterparts. Again, while the variability was substantial, on average a case cost $250 more per month ($366 median). Thirty-nine of the 61 pairs (64%) involved cases that were more expensive than the matched control.

Maximum monthly costs

Table 2 looks at the worst case scenario for each matched pair by using the maximum monthly cost incurred for each patient. The purpose of this approach was to examine the impact, if any, of catastrophic charges. This analysis was also carried out to test the theory that patients who go on trial experience bolus amounts of treatments upon initial entry on study and/or cause the system to incur greater catastrophic costs due to closer monitoring. Cases were slightly higher on average ($177 and $1,342 mean and median respectively), but there were almost as many of the 61 pairs that had the control patient's worst cost being higher than the case patient (25 or 41%) than vice versa (36 or 59%).

Costs from index date (TEED)

Costs incurred for varying periods from the index date are summarized in figure 6. In the first month of being on trial the study patients cost an average of $569 more than those who did not go on trials. Costs incurred over the first three months were almost identical and appear to be relatively high compared to months further away from the TEED. The cost increase of being on trial remained consistently within the 5-11% range in terms of the average cost for those who did not go on trials. Figure 6 indicates the stark differences between the mean and median cost estimates due to the extreme values observed in figures 1 and 2. Differences beyond the second year on study become meaningless as the number of patients surviving beyond that point becomes small.

Costs from date of death

Costs incurred in the months preceding death are contained in figure 5. The average difference is fairly consistent at around $1,000 a month, with the cases incurring greater costs. Roughly 65% of the 34 pairs where death occurred in both patients within the matched pair were such that the cases incurred greater costs consistently over the last year of life.

The previous total costs averaged around $42,000 per control. Costs incurred in the last month of life represent roughly 7% of the total. Costs for the last three months of life amount to roughly 17% of the total for the study period. There was no evidence to suggest that costs incurred in the final month or three months of life amount to a disproportionately high portion of the total health care costs incurred during the study period.

Survival-related analysis

Figure 3 presents the usual Kaplan-Meier survival curves for the cases and controls. The patients who did not go onto a clinical trial had a slightly better survival profile than those who participated in a clinical trial. Follow-up was roughly one month shorter for patients that went on clinical trials (17.6 versus 18.4 on average). Kaplan-Meier survival curve estimates indicated that this difference was not statistically significant (logrank p-value=0.06). This presented us with a potential concomitant confounding variable for examining the relative costs for the two patient groups. Figure 7 presents these data in another manner, organising the data within a butterfly plot into those who died within the five-year follow-up period.

Figure 4 displays the pseudo-survival curves for patients on trial versus those who did not go on trial using total costs as the X-axis variable and the proportion of patients that had total costs equal to or below each given level of cost. This approach is similar to examining the survival curves for two treatment groups. This approach must be used with caution. However, it has been shown that the application of survival methods for cost data does produce biased estimators. In our context, the approach provides a reasonable supplementary descriptive tool which confirmed results of the other approaches.

The KMSA estimators, which incorporate the relative likelihood of survival as a weighting function for the cost data, provided further evidence that the differences were slight. The average cost for the cases was $1,700 compared to $2,220 for the controls (Figure 8).

The mathematical details of the KMSA approach has as of yet to be worked out and so it is not possible to produce the exact nature of the distribution for the KMSA estimator. We bootstrapped 10,000 samples from the original data to obtain estimates of variability for the KMSA estimator. The average total cost for patients on trial was $46,563 with a standard deviation of $6,807 while the mean and standard deviation for those who did not go on trial were $43,820 and $8,484 respectively. These results all pointed to the conclusion for the incremental costs of a patient's participation in a clinical trial as having a relatively small impact on the total health care cost for an individual.

SUMMARY

The analysis of health care cost data presents some unique challenges that require more than routine statistical analysis available in the present array of SAS procedures. Basic methods such as normalizing transformations and adjusting for inflationary influences are useful but do not solve all of the problems. Defining the endpoint can in fact be the most difficult part of the analytical process.

The novel analytical methods have been introduced so recently that code for carrying out the procedures was not available in any computer language. We hope that through the availability of this series of SAS macros that the novel analytical methods can be further explored as well as facilitating some of the more established statistical procedures. It is comforting to note that despite the varied and sundry statistical procedures applied to our cost data, the results remained remarkably consistent across all methods be they descriptive or inferential. SAS code may be obtained from the authors at jsloan@mayo.edu.

Acknowledgements

The authors are indebted to Dr. Ruth Etzioni for generously sharing her expertise on the KMSA methodology and for putting us back on track whenever we went astray.

References

Adams ME McCall NT Gray DT Orza MJ Chalmers TC. Economic analysis in randomised control trials. Medical Care 30: 231-243, 1992.

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

Etzioni R Feuer EJ Lin D Sullivan SD Ramsey SD. On the use of survival analysis techniques to estimate medical care costs. (unpublished manuscript under review, 1997).

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

Grieve AP. Issues for statisticians in pharmaco-economic evaluations. Statistics in Medicine 17: 1715-1723, 1998.

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

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

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

Wagner JL Alberts SR Sloan JA Cha S et al. The incremental costs of enrolling cancer patients in clinical trials: a population-based study. Journal of the National Cnacer Institute , 1999 (under review).

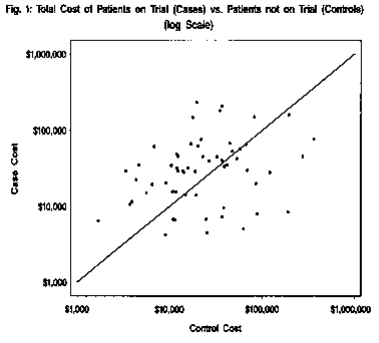
Zhou X Melfi CA Hui SL. Methods for comparison of cost data. Annals of 127: 752-756, 1997.

Table 1: Total Cost Summary Statistics

Statistic	On Trial Cases (n=61)		No Trial Controls (n=61)		Intra-Pair Difference (Case – Control)		
	Raw	Discount	Raw	Discount	Raw	Discount	P-value
Total costs from index date							
Total	2.7 M	2.6 M	2.6 M	2.5 M	150,466	140,050	
Mean	44,555	42,967	42,089	40,671	2,467	2,296	0.80
Std	50,986	48,556	64,358	62,423	74,354	71,457	
Median	29,833	29,539	19,395	18,831	7,284	7,081	0.13
Minimum	545	543	614	612	-282,771	-269,217	
Maximum	234,763	227,227	359,981	346,022	215,018	208,120	
Costs per month of follow-up							
Total	154,724	151,932	139,684	137,126	15,040	14,806	
Mean	2,536	2,491	2,290	2,248	247	243	0.61
Std	2,506	2,494	3,632	3,593	3,808	3,764	
Median	2,052	1,951	1,100	1,089	366	359	0.06
Minimum	89	86	63	59	-17,077	-16,969	
Maximum	15,319	15,305	22,751	22,605	12,838	12,835	

Table 2: Maximum monthly costs per patient

Statistic	On Trial Cases (n=61)		No Trial Controls (n=61)		Intra-Pair Difference (Case – Control)		
	Raw	Discount	Raw	Discount	Raw	Discount	P-value
Mean	10,709	10,361	10,531	10,216	177	144	0.95
Std	12,492	12,125	16,407	16,130	20,583	20,093	
Median	6,379	6,265	5,545	5,396	1,342	1,299	0.36
Minimum	278	276	268	268	-73,560	-73,665	
Maximum	72,178	72,178	82,095	82,095	68,003	68,115	



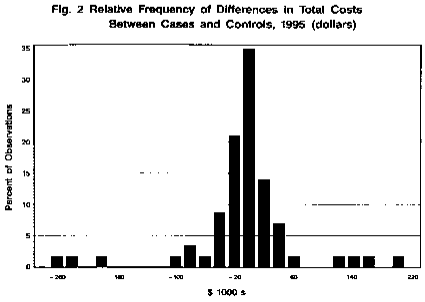


Fig. 3 Survival in Months from TEED for Cases and Controls

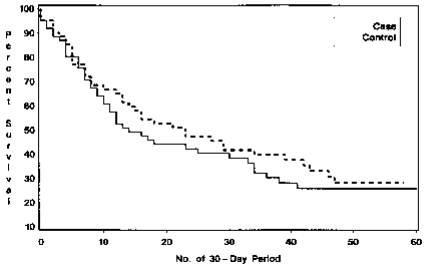


Fig. 4 Using Cost as a Time Variable to Compare Cases and Controls

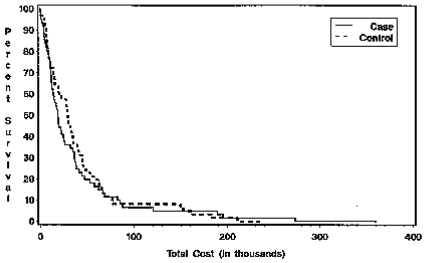
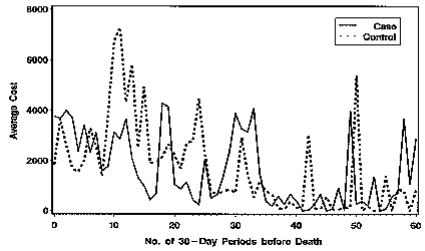


Fig. 5 Average Monthly Cost from Death for Cases and Controls

Mean



Median

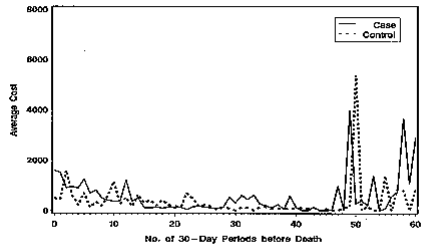
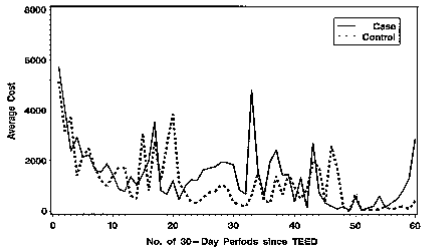


Fig. 6 Average Monthly Cost from TEED for Cases and Controls

Mean



Median

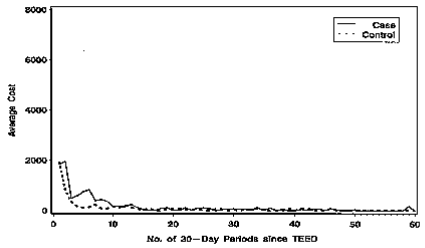


Fig. 7: FU Information (Case & Control)

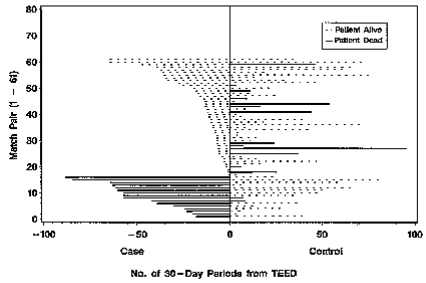


Fig. 8 Cumulative Costs (KMSA Method) of First-Stage Matches for Cases and Controls