

A Comparison of Analytic Models for the Costs of the Hospitalized Diabetic Patients

Giulia Zigon,¹ Rosalba Rosato,² Simona Bò³ and Dario Gregori⁴

Abstract

OBJECTIVE: To compare some survival models as applied to the estimation of the costs of hospitalization as a function of several covariates in diabetic patients. The application of the Aalen regression model (Aalen, 1989) to medical costs is stressed. **DESIGN:** Retrospective observational study analyzing hospitalizations in a cohort of diabetic patients with a follow up of 4.5 years. **STUDY POPULATION:** A total of 2550 patients have been included in the cohort, according to clinical-based enrollment standards. The patients with at least one hospitalization have been considered in this analysis. **METHODS:** Costs have been modelled using five different regression model: the ordinary least square regression model, the logistic regression model using the median and the third quartile of the costs distribution as cut-off points, the parametric survival model assuming Weibull distribution, the Cox proportional hazard model and the Aalen additive regression model. **CONCLUSIONS:** The Aalen additive regression model applied to the costs has the best performances in estimating the mean hospitalization costs for specific clinical profiles.

1 Introduction

Type 2 diabetes is a common disorder, affecting approximately 100 million people worldwide. An increase of prevalence of about 50% is also expected in the next years for Europe (Henriksson, 2000). This increase may be attributable to several factors such as longer survival and also changes in life style such as low physical activity, obesity, high caloric in-take and smoking.

Diabetes is characterized by the appearance of acute and chronic complications, that determine an elevated risk of hospitalization. The most common diabetes complications are foot ulcers, eye and limb disease, heart disease, retinopathy, neuropathy and nephropathy (Leese, 1992; Reiber, 1992; Shobhana, 2002; Vijan, 2000). It has been reported that diabetic patients are more than twice as likely to be hospitalized as the population as a whole. The average length of hospital stays for diabetic patients are significantly greater than non-diabetics patients (Carral et al., 2002), and this obviously has considerable consequences on the total cost of care and thereafter on the organization of public services.

¹Department of Statistics- viale Morgagni, 59- University of Firenze, Italy

²Unit of Cancer Epidemiology, University of Torino, Italy

³Department of Internal Medicine, University of Torino, Italy

⁴Department of Public Health and Microbiology, University of Torino, Italy

Costs related to diabetes and diabetic complications are two or three times higher than those for patients not affected by diabetes (Rubin et al., 1994). The reason for this may be associated with co-morbidities, more frequent in these patients, disease evolution, poor compliances with diet, and pharmacological treatment. The direct economic costs of diabetes are for the most part medical charges (hospitalizations, home visits, laboratory tests, pharmaceutical expenditure, and other medical and paramedical services), whereas the indirect costs encompass loss of working days and a range of effects on the patient's quality of life (Triomphe, 1991).

The estimation of the total direct cost is not an easy task, in particular when the goal of the analysis is to relate costs to specific pattern of covariates. Most studies focus on estimating the costs of care over a pre-specified time period, dealing with the early deaths as censored observations. Indeed, accurate cost estimation is problematic when cost records are incomplete, because censoring could lead to biased estimates of costs unless it is appropriately accounted in the analysis (Dudley et al., 1983; Fenn et al., 1995, 1996). With reference to these characteristics and particularly to the presence of censoring, several works in literature (Dudley et al., 1983) have proposed to use the survival models as the Kaplan Meyer and the Cox regression model. It is possible use the survival models, because, from a mathematical point of view, accumulated costs behave pretty much like times, being a continuous non-decreasing variable ($x_i \geq x_{i-1}$). However, assumptions behind the survival model are often violated in cost estimations.

The problems of cost analysis regard:

- the possible large mass of observations with zero cost;
- the asymmetry of the distribution, given that there is a minority of individuals with high medical cost compared to the rest of the population;
- the censoring must be independent or non informative. This condition is needed because the individuals still under observation must be representative of the population at risk in each group, otherwise the observed failure rate in each group will be biased and it is hardly satisfied in a cost analysis;
- the assumption of proportional hazards may be violated by the medical costs; in general, the proportionality will not be satisfied when costs are accumulated at different rates (Etzioni et al., 1999)

The aim of this study is modelling the hospitalizations costs up to 4.5 years from the enrollment in the cohort. In this case, deaths have to be considered as censored data, since non surviving patients would likely have had higher costs if they had survived longer. An analysis that does not account for this possibility may bias the assessment of the importance of the severity of illness on the final costs. Moreover, the second goal is to describe the behavior of costs for diabetic patients as a function of demographic and clinical covariates, chosen among those considered as classical risk factors for diabetes (Brandle et al., 2003).

We compared the performances in accomplishing this task of several statistical models, all used in the medical literature routinely in this setting: the ordinary least square (OLS) regression model, the logistic regression model using the median and the third

quartile of the cost distribution as cut-off points, the parametric survival model assuming Weibull distribution, the Cox proportional hazard (PH) model. In addition, we introduce the use of the Aalen additive regression model for modeling costs (Aalen, 1989; Aalen, 1993).

2 Methodology

2.1 The data

This cost analysis is based on the data arising from a retrospective observational study analyzing the repeated hospitalizations in a cohort of diabetic patients.

A cohort of 3892 subjects, including all type 2 diabetic patients, resident in region Piedmont, attending the Diabetic Clinic of the San Giovanni Battista Hospital of the city of Torino (region Piedmont, Italy) during 1995 and alive at 1st January 1996 was identified. A mortality and hospitalization follow-up was carried over up to 30th June 2000. The patients were included in the study if they had at least one hospitalization in the subsequent years of follow-up (selecting only more severe patients is aimed at reducing the sample selection bias and making the cases more homogeneous) and according to clinical based enrollment standards: age at diagnosis, therapy type and glucagon test. These criteria lead to the definition of the final dataset for the analysis, based on 2550 patients. A total of 4816 ordinary hospitalizations and 2183 daily hospitalizations have been observed during the 4.5 years follow-up and about 13% of the sample (342 patients) died during the follow-up. Demographic data (age, sex) and clinical data relative to the year 1995 (duration of disease or years of diabetes and number of other co-morbidities) were utilized for the present analysis. The costs (in euros) for the daily and the ordinary hospitalizations have been calculated referring to the Italian DRG (diagnosis related groups) system.

2.2 The models

Several models were considered and contrasted for the analysis. Most of them are relying on very diverse hypotheses and tend to have a different statistical motivation.

Ordinary Regression The OLS model assumes the following form for the costs

$$c_i = \sum_{j=1}^h \beta_0 + \beta_j x_{ij} \quad (2.1)$$

where c_i ($i = 1, \dots, k$) are the ($k = 2550$) sums of the hospitalizations costs calculated for every patient, and x_{ij} are the j ($j = 1, \dots, h$) covariates, β_0 is the intercept and β_j are the j corresponding regression coefficients estimated with the least square method. In this model the censored costs are treated like the complete costs.

Logistic Regression The logistic model has been utilized to model the probability of having costs in excess of a given threshold. In our setting, we used two different cut-off

points in the costs distribution: the median q_2 and the third quartile q_3 (the choice of these cut-off points is only for illustrative purposes).

$$p(c_i > q_{2,3}) = \frac{1}{1 + \exp(-\sum_{j=1}^h \beta_j x_{ij})} \quad (2.2)$$

Thus, we estimated the probability to have a cost greater than the median and the third quartile in the logistic regression model, as a function of the covariates (considering the censored observations as complete observations).

Survival models Using the survival models in the cost analysis, we define the cost function as $S(c) = p(c_i > c)$, where c_1, \dots, c_k are the costs for every subject, that is the probability that the dependent variable cost is greater than a certain cost c ; the hazard rate function $\lambda(c)$ is defined as $\lambda(c) = f(c)/(1 - F(c))$ where $F(c) = 1 - S(c)$ and $f(c)$ is the p.d.f and $F(c)$ is the c.d.f.

The parametric survival model assuming Weibull distribution and the Cox PH regression model relate the hazard at each cumulative cost c_i to the covariates:

$$\lambda(c_i | \underline{x}_h) = \lambda_0(c) \exp\left(\sum_{j=1}^h \beta_j x_{ij}\right) \quad (2.3)$$

where $\lambda(c_i | \underline{x}_h)$ is the hazard rate of a cost c_i for an individual i with a covariates vector \underline{x}_h and $\lambda_0(c)$ is the base line risk function. In the Weibull regression model the base line $\lambda_0(c)$ is assumed distributed in according to the Weibull distribution and in the Cox model no assumption are made about the base line function.

Aalen model The Aalen model (Aalen, 1989; Aalen, 1993) assumes the variables interact in an additive manner on the risk function, and estimates the increase of risk associated with the covariates in absolute terms. The Aalen model can be expressed in the following fashion:

$$\lambda(c_i | \underline{x}_h) = \lambda_0 + \sum_{j=1}^h \alpha_j(c) x_{ij}(c) \quad (2.4)$$

where the hazard rate $\lambda(c_i | \underline{x}_h)$ is a linear combination of the variables $x_j(c)$ and $\alpha_j(c)$ are h functions estimated from the data. The slope of the h cumulative regression functions demonstrates the weight of each covariate on the hazard function, while the costs are on the x-axis, and in the case in which the covariate has no effect the function should be a straight line near 0. The Aalen model can also be re-written in other terms considering the observations censored:

$$\underline{\lambda}(c) = \underline{Y}(c)\underline{\alpha}(c) \quad (2.5)$$

where $\underline{\lambda}(c)$ is a k hazard vector, $\underline{\alpha}(c)$ is a h functions vector and $\underline{Y}(c)$ is a matrix with k rows corresponding to the individuals and h columns corresponding to the h covariates including the constant. This matrix is constructed in such a way that if the i -th individual is still in the risk set, then the i -th row will be formed by the vector of their covariates,

while if the individual is no longer in the risk set the corresponding row will count only 0. In the case of a censored observation, the i -th row will only count zeros.

Starting from this expression of the model the estimation of the h cumulative regression functions vector $\underline{\alpha}(c)$ is obtained with the last square method using the generalized inverse of the matrix $\underline{Y}(c)$.

2.3 The mean costs

Five clinical profiles have been considered, with different values for the variables Age, Sex, Years of diabetes and Number of co-morbidities. The predicted mean costs have been calculated for each model except for the logistic model; in this case, the probabilities to have costs greater than the cut-off points have been estimated for the logistic model.

For the PH and the Aalen model the mean costs have been calculated considering the area (calculated with numeric integration) under the estimated survival curve; the median value is the cost corresponding to the value 0.5 in the estimated survival curve.

3 Results

The total hospitalizations costs after 4.5 years of follow-up are characterized by an asymmetric distribution (Figure 1); there is a considerable presence of censored data, due to the patients' deaths (342 patients, about 13%).

Table 1: Cost distribution in Euro (mean and quartiles).

Mean	Min	Iq	IIq (Median)	IIIq	Max
7278	99.42	1938	3913	9014	89650

The costs (First quartile, Median, Third quartile) for each level of the considered variables are reported in Table 2. Classes for the covariates Years of diabetes and Age have been constructed according to their quartiles.

Several crucial assumptions are not met for several models. The normality assumption about the residuals for the OLS model is not satisfied (Shapiro-Wilk test $p < 0.001$). The assumption of proportional hazard is not satisfied, as proven by the Grambsch-Therneau (Grambsch et al., 1994) χ^2 equal to 40.88, $p < 0.001$. The five regression models applied to the diabetic data using Age (as continuous variable), Sex, number of Years of diabetes (as continuous variable) and number of Co-morbidities (as continuous variable) give significant effects for all the covariates except Years of diabetes in both the Logistic models (Table 3). The cumulative regression functions estimated for the Aalen model are presented in the Figure 2.

An index for the explained variation has been calculated for each model. The Dxy rank correlation (Schemper et al., 1996) is used for all the models but for the OLS regression model, where the R^2 was used. The logistic model with median (Dxy=0.268) and third quartile (Dxy=0.263) as cut-off points, the Weibull model (Dxy=0.236) and the Cox model (Dxy=-0.231) did not perform well in terms of explained variation; the same occurs for the OLS regression model ($R^2 = 0.033$). The Aalen model on the contrary

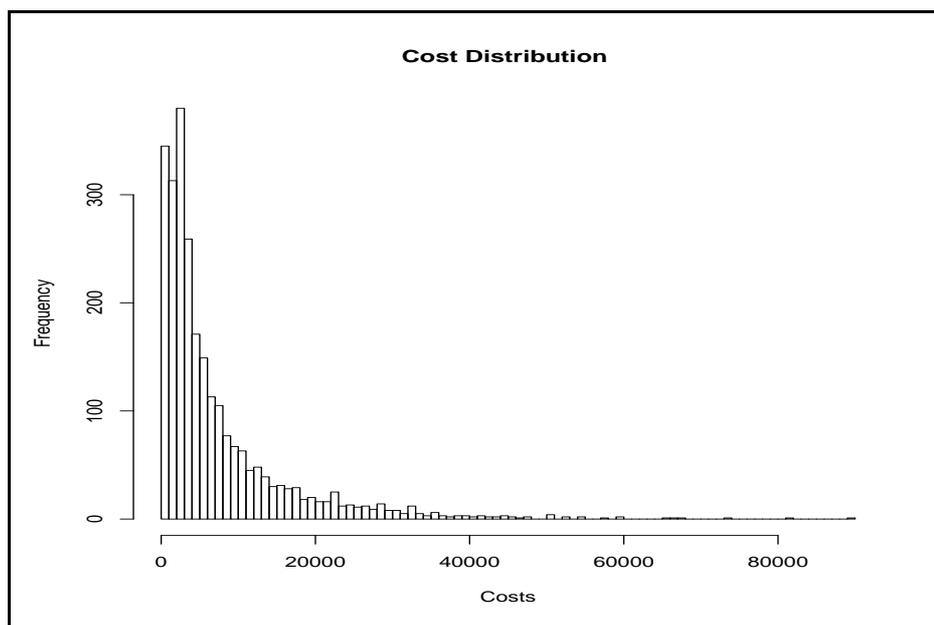


Figure 1: Histogram of total costs in Euro at 4.5 years of follow-up.

gave a reasonable performance ($D_{xy} = -0.978$). The negative sign of the index is due to the increment in the estimated survival corresponding to a decrement in the area under the curve (costs). The overall mean (with a 95% confidence interval) and the overall median costs estimated for each model are presented in Table 4 (the mean and the median probabilities are presented for the logistic models)

The comparison of the estimated costs for different clinical profiles allows us to have a specific measure of the estimation bias (difference between the observed and the estimated cost) for each model. The estimated mean costs for five different clinical profiles are presented in Table 5. The observed costs in the table refer to the mean of the costs for the patients respectively in the age classes [35, 45), [65, 75), [55, 65), [60, 70) and in the classes [0,4), [8,12), [18,22) [13,17) for the variable Years of diabetes.

The Logistic regression model predicts a probability equal to $p_m = 0.2667$ to have a cost greater than the median (3913 euros) and equal to $p_{3q} = 0.1355$ to have a cost greater than the third quartile (9014 euros) for a forty-year-old female diabetic for 2 years, $p_m = 0.3961$, $p_{3q} = 0.2452$ for a forty-year-old female diabetic for 10 years with another co-morbidity, $p_m = 0.6891$, $p_{3q} = 0.4234$ for a seventy-year-old men diabetic for 20 years with another co-morbidity, $p_m = 0.5371$, $p_{3q} = 0.3218$ for a sixty-year-old female diabetic for 15 years with another co-morbidity and $p_m = 0.5457$, $p_{3q} = 0.2646$ for a sixty five-year-old men diabetic for 30 years.

4 Discussion

The data about the costs are not normally distributed because a minority of individuals accumulates disproportionately high medical care costs compared to the rest of the population; in addition, there is the problem of the incompleteness of the follow-up data due to

Table 2: Distribution of costs (I quartile, Median, III quartile) according to the covariates.

		N	(Iq, Me, IIIq)
Sex	Female	1270	(1872, 3617, 8424)
	Male	1280	(2047, 4290, 9700)
Co-morbidities ≥ 1	No	2187	(1850, 3704, 8386)
	Yes	363	(2765, 5943, 12950)
Years of Diabetes	[0, 4)	480	(1641, 3552, 8452)
	[4, 10)	594	(1922, 3728, 8009)
	[10, 18)	691	(1886, 4007, 9363)
	[18, 48]	785	(2142, 4307, 9671)
Age	[22.1, 59.2)	638	(1425, 2891, 7261)
	[59.2, 66.2)	638	(1872, 3684, 8121)
	[66.2, 72.6)	637	(2395, 4844, 10940)
	[72.6, 90.8]	637	(2333 4517, 9411)
Overall		2550	(1938, 3913, 9014)

Table 3: Coefficients estimated in the models with standard errors (significant coefficients in cursive).

Models	β_1 Intercept	β_2 Age	β_3 Sex	β_4 Years Diabetes	β_5 N. co- morbidities
OLS	<i>2155.15</i> (SE=1220.02)	<i>53.70</i> (SE=18.50)	<i>829.80</i> (SE=360.65)	<i>59.02</i> (SE=21.36)	<i>2946.98</i> (SE=474.93)
Logistic (median)	<i>-2.102</i> (SE=0.283)	<i>0.026</i> (SE=0.004)	<i>0.346</i> (SE=0.081)	<i>0.006</i> (SE=0.004)	<i>0.539</i> (SE=0.110)
Logistic (III quartile)	<i>-2.565</i> (SE=0.330)	<i>0.017</i> (SE=0.004)	<i>0.233</i> (SE=0.093)	<i>0.005</i> (SE=0.005)	<i>0.682</i> (SE=0.111)
Weibull	<i>3.0683</i> (SE=0.348)	<i>0.0577</i> (SE=0.005)	<i>0.2032</i> (SE=0.107)	<i>0.0439</i> (SE=0.006)	<i>1.3073</i> (SE=0.160)
Cox	– –	<i>-0.0196</i> (SE=0.001)	<i>-0.0938</i> (SE=0.03)	<i>-0.0149</i> (SE=0.001)	<i>-0.4829</i> (SE=0.051)
Aalen	<i>4.611</i> (SE=5.744)	<i>0.023</i> (SE=0.067)	<i>0.873</i> (SE=1.503)	<i>-0.078</i> (SE=0.118)	<i>-1.504</i> (SE=0.576)

the early deaths. Taking in consideration such characteristics of the costs, several works in the literature have been recently proposed (Bang and Tsiatis, 2000; Lin et al., 1997) to deal with these issues. However, most of these methods rely to iterative procedures for median regression, which are still presenting difficulties in obtaining stable and easy-to-converge estimates (Bang and Tsiatis, 2000). Standard regression procedures, basically based on Cox proportional hazard model has been heavily criticized, due to severe lacks regarding underlying assumptions of such models. Among all issues, the major point raised, is the effect of a non-constant accumulation of costs over time, which induces a non proportionality of hazards in the Cox model (Etzioni et al. 1999).

To overcome such problems, we proposed to use the Aalen regression model (Aalen,

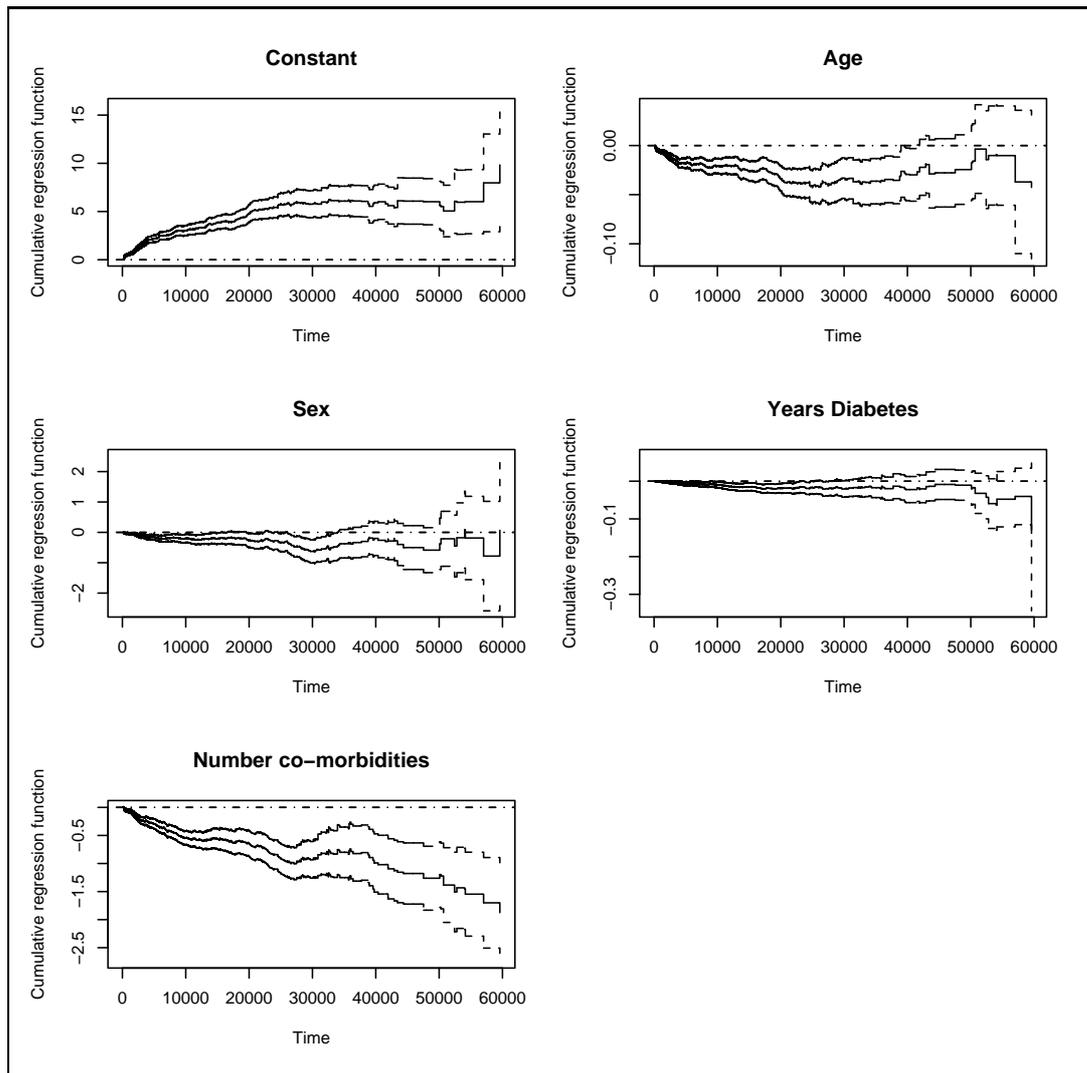


Figure 2: Plot of Aalen coefficients as a function of costs. Negative hazard implies an increment in costs for the specific covariate.

1989) which is robust for the non proportionality in hazard and also for the change in number and measurement units of the covariates. One disadvantage of such model is of computational nature, because for the more extreme observations the number of covariates that can be used is limited, to avoid that the Y matrix loses its full rank. To overcome this computational issue for the highest costs, it may be needed to have a large number of observations also for the more extreme costs, which is generally difficult.

In this analysis we compared the results in the costs estimations for five different regression models. The aim was to find out what is the regression model, which provides the best estimations of the overall mean and median costs and also the predicted costs for patients with specific clinical profiles. The comparison of the models was based on the differences between the real observed costs and the estimated costs.

We found an increase in the risk of high hospitalization costs directly connected to the considered variables: age, sex, duration of disease and presence of other pathologies.

Table 4: Median and mean costs estimated for each model.

Models	Median Cost	Mean cost (95% C.I.)
OLS	7013	7278 (7222.88 7333.12)
Logistic (Median)	P=0.4962	P= 0.500 (0.480 0.519)
Logistic (III quartile)	P=0.2352	P=0.2502 (0.2334 0.2670)
Weibull	7678	8269 (8154.698 8383.302)
Cox	4344	8717.984 (7881.01 9554.95)
Aalen	4879	8077.735 (7493.737 8661.733)

Table 5: Mean costs estimated for specific clinical profiles. Numbers in () are referring to 95% confidence interval.

Observed costs	OLS (95% C.I.)	Weibull (95% C.I.)	Cox (95% C.I.)	Aalen (95% C.I.)
	Age=40	Years of Diabetes=2	Sex=F	Co-morbidities=0
3388	4421 (4365.88 4476.12)	236.40 (102.3689 370.4311)	1517.058 (1229.894 1804.222)	3936.722 (3272.815 4600.629)
	Age=40	Years of Diabetes=10	Sex=F	Co-morbidities =1
7894	7840 (7784.88 7895.12)	1242 (1107.969 1376.031)	4594.555 (3521.434 5667.676)	5108.192 (4043.500 6172.884)
	Age=70	Years of Diabetes=20	Sex=M	Co-morbidities=1
8077.704	10870 (10814.88 10925.12)	13347 (13212.97 13481.03)	16401.33 (13488.79 19313.88)	7637.626 (6401.272 8873.980)
	Age=60	Years of Diabetes=15	Sex=F	Co-morbidities=1
5724.294	9209 (9153.88 9264.12)	4909 (4774.969 5043.031)	9806.214 (8006.951 11605.477)	6411.435 (5374.243 7448.626)
	Age=65	Years of Diabetes=30	Sex=M	Co-morbidities=0
5527.482	8246 (8190.88 8301.12)	4199 (4064.969 4333.031)	8835.986 (7363.122 10308.850)	5377.089 (4574.917 6179.260)

There is agreement in all the models about the effects of the considered covariates, (except the variable duration of disease in the Logistic models). In particular the presence of co-morbidities has the biggest effect on the costs, while the sex is not a very important discrimination factor in the cost expenditure. The logistic regression is not very informative in this analysis, but the estimated probabilities to have a cost greater than the median (estimated probability $p=0.50$) or the third quartile (estimated probability=0.2502) are accurate. The odds ratios in the logistic models confirm the weak effect of the variable sex: $OR = 1.41$ (median as cut-off point) and $OR = 1.26$ (third quartile as cut-off point), and the considerable effect of the variable Number of co-morbidities: $OR = 2.94$ (median as cut-off point) and $OR = 3.92$ (third quartile as cut-off point). The probability to accrue high hospitalization costs (costs greater than 3913 euros) is almost three times and almost four times (for total costs greater than 9014 euros) for the patients, that present another co-morbidity with the diabetes respect the patients only diabetic.

The Aalen model gives the best overall mean value (after the unbiased estimation of the OLS model) with a difference of 799.235 euros with respect to the mean value (11%). Nevertheless it overestimates the overall mean cost like the considered Cox model (a difference of 1439.984 euros accounting for about 20%) and the Weibull model (a difference of 991 euros, about 14%).

In the Cox model there is a tendency to overestimation for the high risk profiles (third,

fourth and fifth clinical profile), while there is a tendency to underestimate the costs for the low risk profiles (first and second). In the Weibull model there is a general tendency to underestimate the costs for the considered clinical profiles and particularly for the low risk profiles (first and second profile), while the highest risk profile (the third: a seventy-year-old men, diabetic for 20 years and with another co-morbidity) is strongly overestimated (observed cost: 8077.7904 euros, estimated mean cost: 13347 euros). The Aalen model gives the best estimations of the mean costs in all the considered profiles (except for the second profile, where the OLS gives the best estimation). The bias in the estimations is included between the 2.72% (for the fifth profile) and the 16.18% (for the first profile).

In view of our results the Aalen model shows to be a realistic and informative way to characterize the effect of covariates on costs at a pre-specified time (in our case, 4.5 years). However, a key assumption of this model remains the assumption of independent or non informative censoring, which means that censored individuals cannot constitute a particularly high or low risk subgroup; rather they should be representative in terms of their risk of hospitalization cost. We consider in this study the deaths as censored data, because the aim is the estimation of the total costs at a fixed time of 4.5 years (total follow up time). The censoring data constitutes about 13% of the sample in this study and the high-risk individuals tend to be censored (they die) more likely than those remaining and thus they will constitute a selective, low-risk sample, and the estimated hazard could possibly lead to an underestimation of the true cost hazard. It has to be noticed however that such effect is common for the vast majority of the cost models presented in the literature (Lin et al., 1997) and thus it remains a major issue for forthcoming research.

Acknowledgement

This project has been conducted under a grant of Regione Piemonte (Italy), "Ricerca Sanitaria Finalizzata 2003".

References

- [1] Aalen, O.O. (1989): A linear regression model for the analysis of life times. *Statistics in Medicine*, **8**, 907-925.
- [2] Aalen, O.O. (1993): Further results on the parametric linear regression model in survival analysis. *Statistics in Medicine*, **12**, 1569-1588
- [3] Bang, H. and Tsiatis, A.A. (2000): Estimating medical cost with censored data. *Biometrika*, **87**, 329-343.
- [4] Brandle, M., Zhou, H., Smith, B.R., Marriott, D., Burke, R., Tabaei, B.P., Brown, M.B., and Herman W.H. (2003): The direct medical cost of type 2 diabetes. *Diabetes Care*, **26**, 2300-2304.
- [5] Carral, F., Oliveira, G., Salas, J., Garcia, L., Sillero, A., and Aguilar, M. (2002): Care resource utilization and direct costs incurred by people with diabetes in a Spanish hospital. *Diabetes Res Clin Pract.* **56**, 27-34.

-
- [6] Dudley, R.A., Harrell, F.E., Smith, L.R., Mark, D.B., Califf, R.M., Pryor, D.B., Glower, D., Lipscomb, J., and Hlatky, M. (1983): Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *Journal of Clinical Epidemiology*, **46**, 261-273.
- [7] Etzioni, R.D., Feuer, E.J., Sullivan, S., Lin, D., Hu, C., and Ramsey, S. (1999): On the use of survival analysis techniques to estimate medical care costs, *Journal of Health Economics*, **18**, 365-380.
- [8] Fenn, P., McGuire, A., Phillips, V., Backhouse, M., and Jones, D. (1995): The analysis of censored treatment cost data in economic evaluation. *Medical Care*, **33**, 851-863.
- [9] Fenn, P., McGuire, A., Backhouse, M., and Jones, D. (1996): Modelling programme costs in economic evaluation. *Journal of Health Economics*, **15**, 115-125.
- [10] Grambsch, P. and Therneau, T. (1994): Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515-526.
- [11] Henriksson, F., Agardh, C.D., Berne, C., Bolinder, J., and Jonsson, B. (1999): Economics analysis of diabetes is necessary. It facilitates decision-making and international comparison. *Lakartidningen*, **96**, 3915-3919.
- [12] Leese, B. (1992): The costs of diabetes and its complications. *Soc Sci Med.*, **35**, 1303-1310
- [13] Lin, D.Y., Feuer, E.J., Etzioni, R., and Wax, Y. (1997): Estimating medical costs from incomplete follow up data, *Biometrics*, **53**, 419-434.
- [14] Reiber, G.E. (1992): Diabetic foot care. Financial implications and practice guidelines. *Diabetes Care*, **15**, 29-31.
- [15] Rubin, R.J., Altman, W.M., and Mendelson, D.N. (1994): Health care expenditures for people with diabetes mellitus. *Journal of Clinical Endocrinology and Metabolism*, **78**, 809A-F
- [16] Schemper, M. and Stare, J. (1996): Explained variation in survival analysis. *Statistics in Medicine*, **15**, 1999-2012.
- [17] Shobhana, R., Rao, P.R., Lavanya, A., Vijay, V., and Ramachandran, A. (2001): Foot care economics cost burden to diabetic patients with foot complications: a study from southern India. *J Assoc Physicians India*, **49**, 530-533.
- [18] Triomphe, A. (1991): The socio-economic cost of diabetic complications in France. *Diabet Med.* **8**, S30-2.
- [19] Vijan, S, Hofer, T.P., Hayward, R.A. (2000): Cost-utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *Journal of American Statistical Association*, **283**, 889-896.