Soz Präventivmed 1993; *38*: 77–82

0303-8408/93/020077-06 $1.50 + 0.20/0

Exploring models for the length of stay distribution

Christiane Ruffieux, Alfio Marazzi, Fred Paccaud

Institut universitaire de médecine sociale et préventive, Lausanne

Casemix methods are intended to provide adequate comparisons between hospitals or hospital systems[1]. When using casemix methods based on consumption/demand criteria (like Diagnosis Related Groups) one has to deal with discrepant observations or "outliers". These are observations whose consumption/demand component (e.g. length of stay) is far removed from the pattern set by the majority of the data. It is well known that outliers have a great influence on the value of many common statistical summaries (e.g. length of stay means) and comparisons. Moreover, the values and the frequency of outliers fluctuate strongly from sample to sample and, therefore, the common summaries can be unreliable.

When using Diagnosis Related Groups (DRGs), a common practice is to trim (large) length of stay (LOS) values according to two basic strategies: (a) Fixed trimming (for example, 100 days for all DRGs, or 50 days for surgical and 100 for medical DRGs, or explicit clinical criteria) (b) Statistical trimming. Statistical trimming usually includes two options:

T_1: Parametric trimming based on the log-normal model: Trim stays whose log-LOS is larger than $\hat{\mu} + k\hat{\sigma}$, or smaller than $\hat{\mu} - k\hat{\sigma}$, where $\hat{\mu}$ and $\hat{\sigma}$ are estimates of the mean μ and the standard deviation σ of the underlying normal distribution, and k is a fixed constant.

T_2: Nonparametric trimming: Trim stays whose LOS is larger than $\hat{q}_3 + k(\hat{q}_3 - \hat{q}_1)$, where $\hat{q}_1$ and $\hat{q}_3$ are the first and third quartile of the empirical LOS distribution, and k is a fixed constant.

When appropriate,a parametric model provides an approximate description of the complete set of data by means of very few numbers (e.g. $\hat{\mu}$ and $\hat{\sigma}$). This is a powerful data condensation. Moreover, the approach based on parametric models allow the application of standard probability theory for solving most practical problems, like comparing LOS means of several DRGs in different hospitals, or predicting the number of stays with LOS beyond given bounds for different scenarios. Unfortunately, the description provided by the normal model (on the log-LOS scale) fails to take into account the presence of outliers. For this reason, classical estimates of μ and σ – as well as the inferences based on them – can be considerably distorted by outliers. As a remedy, the trimming rule T_1 has been proposed. But this rule is obviously unreliable if $\hat{\mu}$ and $\hat{\sigma}$ are the classical estimates, i.e. arithmetic mean and empirical standard deviation (see, for example, reference[2]). Moreover, the log-normal model might not be an appropriate description of the LOS pattern for several DRGs.

It turns out that certain aspects of any distribution, like quartiles, can be estimated in a way which is almost unaffected by outliers, without any need to assume a parametric model. For this reason, T_2 has been proposed as a rule for trimming. Not surprisingly, the trimmed proportions associated with rule T_2 are more stable than those associated with rule T_1, when both rules are applied to several DRGs[2]. However, trimming is not a generally well-founded remedy for alleviating outlier effects: it may be meaningful or not, depending on the use of the trimmed data. For example, symmetric trimming followed by the arithmetic mean can be useful to improve the usual estimate of the center of a symmetric distribution with respect to bias and other stability measures. But asymmetric trimming can just wipe out the most important LOS differences in comparisons, and produce more and more similar trimmed means for increasing levels of trimming! In general, the effects of trimming are quite obscure.

Also, it would be wrong to think that the pure nonparametric approach provides a generally satisfactory framework for the analysis and exploitation of LOS data. Non-parametric procedures do not make any assumptions about a low-dimensional class of distributions, but concentrate on *specific aspects* (e.g. specific frequencies or percentiles) of the data set; they do not consider the structure as a whole. Beyond these few main aspects, the description provided by nonparametric procedures can be very complex and not easily amenable to the use of standard probability methods. In the limit, this description could even require the complete empirical distribution function.

In this pilot study, we consider the possibility of using parametric models as descriptions of the LOS distributions for 230 DRGs and three consecutive years. We fit a few simple models that are frequently used in survival analysis (thus extending the normal and log-normal framework) to available LOS data. Our main interest is to explore whether the model adequacy is stable across time. A positive answer to this question would open the way to the application of robust estimation methods, that are expected to provide reliable, efficient and little-biased data condensation, when models are only approximately adequate, and, particularly, in the presence of outliers.

Population and methods

The database was made available by the "Service cantonal de recherche et d'information statistique" of Canton Vaud. The years 1988, 1989 and 1990 have been included, covering 8 non-profit and short-stay hospitals and 92160 stays. A DRG is attributed to each hospital stay according to a local modification of the US grouper[3]. Only DRGs with 20 stays per year or more have been included in the analysis.

Survival analysis models are used for the LOS distribution: an individual LOS is considered as a survival time, end of stay corresponding to death. Although LOS data contain various forms of censoring (e.g. transfer to another hospital) we do not include this information in this initial analysis.

A continuous survival T is usually described by a survivor function $S(t) = \mathrm{Prob}(T > t)$. Equivalently, it can be described by its density $f(t) = -dS(t)/dt$, or by the associated hazard function, $h(t) = f(t)/S(t)$ (see, e.g.[4]). We have $S(t) = \exp(-H(t))$, where $H(t)$ is a primitive of $h(t)$ (called the cumulative hazard function, with $H(0) = 0$). More specifically, the following commonly used distribution models are considered:

The lognormal distribution:

$$f(t) = \varphi\left(\frac{\log(t-\xi) - \alpha}{\sigma}\right),$$

where φ denotes the standard normal density. The hazard function is first increasing, and then decreasing. Usually $\xi = 0$ ($\xi \neq 0$ can be used when the minimal LOS is 1 or more than 1 days).

The Weibull distribution:

$$f(t) = \frac{\alpha}{\sigma}\left(\frac{t-\xi}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{t-\xi}{\sigma}\right)^{\alpha}\right).$$

If $\alpha > 1$, the probability of leaving the hospital is increasing during the stay. If $\alpha < 1$, this probability decreases during the stay. Usually $\xi = 0$ ($\xi \neq 0$ can be used when the minimal LOS is 1 or more than 1 days).

The exponential distribution:

$$f(t) = \frac{1}{\sigma} \exp\left(-\left(\frac{t-\xi}{\sigma}\right)\right).$$

This is a special case of the Weibull distribution, with $\alpha = 1$; the instantaneous probability of leaving does not depend on the length of stay.

Computational procedures for robust estimation in general asymmetric models require complex nonlinear programming. Programs are under development using tools described in[5], but for this pilot study we use a simple shortcut. For each model, there is a transformation $t^* = g(t)$ of the LOS scale and a transformation $H^* = G(H)$ of the cumulative hazard scale such that $H^*(t^*)$ is a straight line. More precisely:

$H^* = \Phi^{-1}(1 - \exp(-H))$, $t^* = \log(t-\xi)$ in the lognormal case,
$H^* = \log(H)$, $t^* = \log(t-\xi)$ in the Weibull case,
$H^* = H$, $t^* = t$ in the exponential case.

This suggests the following procedure:

1. fit a straight line to the empirical hazard function $\hat{H}^*$ as a function of t^*;
2. convert the straight line coefficient estimates to density parameter estimates.

We use a robust estimation procedure described in[5] for computing the straight line coefficients. Although the error structure in the transformed hazard scale is quite obscure, this procedure seems to produce an adequate and resistant fit.

The goodness of fit of the estimated densities is assessed by means of a χ^2 statistic based on LOS deciles. A two-stage procedure is applied as a screening tool:

1. Each model from the repertory is tested for each DRG on three consecutive years. A model is "accepted" as a satisfactory description if the test is not significant at the 1% level.

However, if sample size is large enough, no model will fit the data. Therefore, a less stringent criterion is added:

2. If, for a given DRG, no model can be accepted in stage 1, the distribution with the smallest χ^2 statistics is recorded as the "best possible choice".

This rule allows DRGs to be classified into eight categories as follows:

1. *logalone:* the lognormal model alone fits the observed pattern
2. *weibalone:* the Weibull model alone fits the observed pattern
3. *expalone:* the exponential model alone fits the observed pattern
4. *weibexpo:* both the Weibull and the exponential models fit the observed pattern
5. *logexpo:* both the lognormal and the exponential models fit the observed pattern
6. *logweib:* both the lognormal and the Weibull models fit the observed pattern
7. *all:* all three models fit the observed pattern

Tab. 1. Number of DRGs per classes of distributions.

Log-alone (1)	Weib-alone (2)	Exp-alone (3)	Weib-expo (4)	Log-expo (5)	Log-weib (6)	all (7)	none (8)	total
20	23	3	4	22	23	72	63	230

8. *none:* none of the three distributions fit the observed pattern.

Results

230 DRGs with more than 20 stays per year, covering 87673 stays, have been analyzed. Table 1 shows that 167 DRGs were attributed to one of the 7 classes with a defined model (i.e. the model has been "accepted" in stage 1), while 63 did not agree with any of the three models considered. The lognormal distribution, alone (class 1) or combined (classes 4, 5 and 7) agrees with the majority (137/167) of the empirical LOS distributions. The Weibull distribution does almost as well (classes 2, 4, 6 and 7: 122/167). The exponential distribution performs less well (classes 3, 5 and 7: 101/167) especially alone: only 3 DRGs belongs to the specific class 3, while classes 1 and 2 have 20 and 23 DRGs, respectively.
63 DRGs belong to class 8. It is not surprising that for many DRGs none of the models are satisfactory: on the one hand, we have a limited repertory of models, on the other hand large sample sizes tend to make goodness of fit tests significant. However, for 27 DRGs in class 8, the same model has been recorded as "the best choice" (in stage 2) throughout the three consecutive years.
Table 2 gives for each of the 25 most frequent DRGs the class to which it has been attributed, as well as the "best choices" over three years. The titles of these DRGs can be found in the Appendix. There are 8 cases for which a model can be "accepted". An example is DRG 14, represented in Figure 1, where the stepped lines correspond to the observed frequencies and the continuous lines to the fitted Weibull models (and LOS larger than 150 have been truncated). For 9 DRGs in class 8 (none), the same model could be recorded as the "best choice". An example is DRG 373 represented in Figure 2 together with the fitted lognormal distributions (LOS larger than 60 have been truncated). The "best choices" are different for only 8 DRGs in class 8: an example is DRG 184, see Figure 3, where the exponential fit is shown for the three years (and LOS larger than 60 have been truncated).

Tab. 2. Classification of the 25 most frequent DRGs and model recorded as the "best choice" for each year.

DRG	class	1988	1989	1990
373	none	log	log	log
184	none	exp	exp	log
119	none	log	log	log
281	logalone	log	log	log
371	none	log	log	log
162	none	exp	exp	exp
167	none	exp	exp	exp
355	none	log	log	log
243	weibalone	weib	weib	weib
227	none	log	weib	weib
14	weibexpo	weib	weib	weib
209	none	exp	exp	exp
231	none	exp	log	log
324	all	log	exp	log
183	none	log	log	exp
225	none	log	log	log
89	none	log	log	weib
127	weibalone	weib	weib	weib
32	logexpo	exp	exp	exp
369	none	exp	exp	log
381	none	log	log	log
222	none	log	exp	exp
219	logalone	log	weib	weib
60	none	weib	weib	weib
182	logexpo	log	weib	log

Tab. 3. Some frequent DRGs with lognormal pattern.

DRG	Year	Number of stays	Estimated expect.	Estimated parameters		
				μ	σ	ξ
89	88*	294	17.10	2.65	0.62	0.00
89	89*	270	16.95	2.63	0.63	0.00
89	90	351	16.70	2.64	0.59	0.00
324	88*	173	4.46	1.29	0.64	0.00
324	89*	212	4.83	1.35	0.67	0.00
324	90*	256	4.87	1.35	0.65	0.00
373	88	2395	7.67	2.01	0.25	0.00
373	89	2433	7.46	1.97	0.27	0.00
373	90	2538	7.18	1.93	0.27	0.00

Tab. 4. Some frequent DRGs with Weibull pattern.

DRG	Year	Number of stays	Estimated expect.	Estimated parameters		
				θ	β	ξ
14	88*	268	23.99	23.17	0.93	0.00
14	89*	266	24.47	23.55	0.92	0.00
14	90*	285	19.65	19.82	1.02	0.00
60	88	340	4.78	5.33	3.29	0.00
60	89	326	4.72	5.27	3.19	0.00
60	90*	322	4.50	5.03	3.25	0.00
127	88	412	17.85	19.72	1.47	0.00
127	89*	360	17.23	19.10	1.51	0.00
127	90*	411	16.34	18.01	1.44	0.00

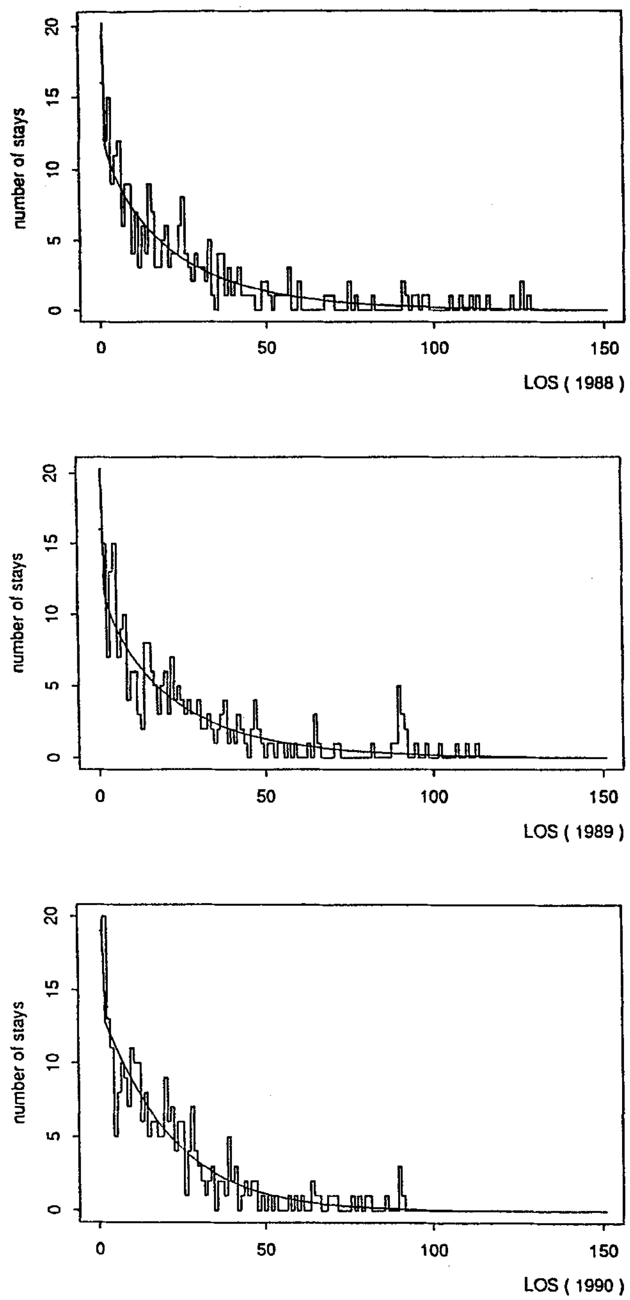

Fig. 1. Weibull model for DRG 14 (Cerebrovascular disorders).

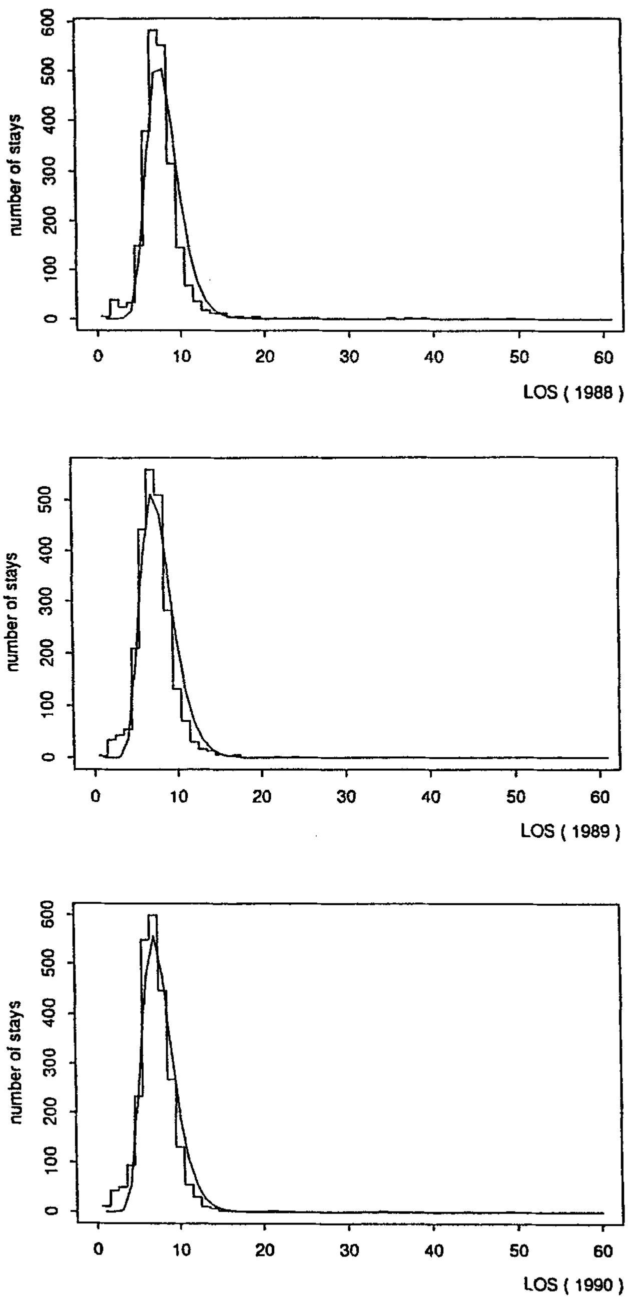

Fig. 2. Lognormal model for DRG 373 (Vaginal delivery).

Tab. 5. Some frequent DRGs with exponential pattern.

DRG	Year	Number of stays	Estimated expect.	Estimated parameters	
				μ	λ
32	88*	255	3.70	1.24	0.41
32	89*	287	3.53	1.40	0.47
32	90*	296	3.43	1.40	0.49
162	88	430	8.88	5.90	0.34
162	89	469	8.52	6.08	0.41
162	90	516	8.13	5.55	0.39
184	88*	296	4.14	1.70	0.41
184	89	377	3.76	1.58	0.46
184	90	384	3.93	1.60	0.43

Tables 3, 4 and 5 show a sample of the most frequent DRGs, the model retained for the three years, the number of stays, the estimated expectation, and the estimated parameter values. A star near the year indicates that the fit is not rejected. It can be seen that, whether the test of goodness of fit is accepted or rejected, the estimated values of the parameters as well as the estimated expectations are quite stable over the years. It appears in most cases that the expectation decreases slightly over the years.

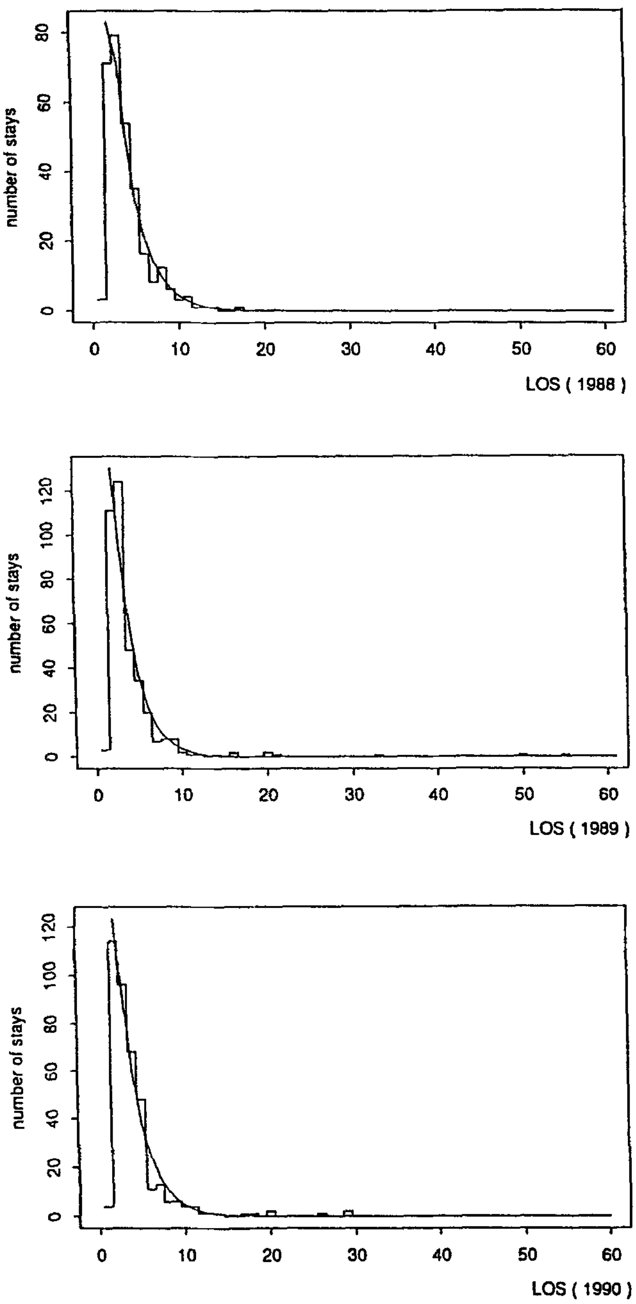

Fig. 3. Exponential model DRG 184 (Esophagitis gastroentesti nal and miscellanous digestive disorders age 0–17).

Conclusion

There is a growing number of opportunities to use DRG databases [6–8]. Moreover, length of stay is directly or indirectly used as a tool for epidemiological or health services surveillance [9–12].
This pilot study explores the possibility of describing DRGs by means of parametric models, and of taking advantage of the parametric approach, which includes a well-founded theory for the study and the analysis of the outlier effects. The results are promising: indeed, it was possible to retain a model as "acceptable" or as "the best choice" over three consecutive years for 194 among the 230 available DRGs.

For 36 DRGs, our screening procedure was not able to assign a satisfactory model from the repertory which has been considered. Clearly, the set of candidate models should be reconsidered and possibly extended. Moreover, appropriate robust procedures for model fitting and for assessing the quality of fit have still to be developed. Finally, the approach has to be compared to the currently used trimming rules. In this perspective, a more extensive study is being planned. The study will use data from different countries in order to assess the geographical stability of the parametric descriptions.

Summary

Diagnosis Related Groups (DRG) are frequently used to standardize the comparison of consumption variables, such as length of stay (LOS). In order to be reliable, this comparison must control for the presence of outliers, i. e. values far removed from the pattern set by the majority of the data. Indeed, outliers can distort the usual statistical summaries, such as means and variances. A common practice is to trim LOS values according to various empirical rules, but there is little theoretical support for choosing between alternative procedures. This pilot study explores the possibility of describing LOS distributions with parametric models which provide the necessary framework for the use of robust methods.

Résumé

Modélisation de la distribution des durées de séjour: Etude exploratoire

Les DRG (Diagnosis Related Groups) sont souvent utilisés pour standardiser des comparaisons entre variables d'utilisation des soins, telles que les durées de séjour (LOS). De telles comparaisons n'ont de sens que si l'on contrôle l'influence des outliers, c'est à dire d'observations éloignées de la majorité des données. La présence d'outliers a pour effet de diminuer la fiabilité des résumés statistiques habituels tels que moyenne arithméthique et variance. Une procédure communément utilisée est le trimming, le point de trimming étant choisi selon différents critères. Mais sans autres hypothèses, on n'a que peu de bases théoriques sur lesquelles s'appuyer pour comparer ces techniques et en choisir une. Le but de cette étude pilote est d'examiner la possibilité d'utiliser des modèles paramétriques pour décrire les distributions de durées de séjour. Ce n'est que dans ce cadre théorique qu'on pourra développer une procédure d'estimation robuste optimale.

Zusammenfassung

Modellisierung der Verteilung der Aufenthaltsdauer: Eine Pilotstudie
Die DRG (Diagnosis Related Groups) werden oft dazu benützt, um den Vergleich von Pflegeleistungen, wie etwa der Hospitalisationsdauer, zu standardisieren. In solchen Vergleichen muss der Einfluss von Ausreissern (d. h. Beobachtungen, die von der Mehrheit der Daten entfernt sind) berücksichtigt werden, da sonst die üblichen statistischen Werte (z. B. Mittelwerte und Abweichungen) unzuverlässig sind. In der Praxis, werden Ausreisser oft aufgrund von empirischen Kriterien ausgeschlossen. Es gibt aber wenige theoretische Grundlagen für solche Verfahren. Ziel dieser Pilotstudie ist die Anwendbarkeit parametrischer Modelle für die Hospitalisationsdauer – und somit die Grundlagen für die Anwendung robuster Verfahren – zu testen.

Appendix

DRG Titles (third revision)

14 Specific cerebrovascular disorders except TIA
32 Concussion age 18–69 without (w/o) complications or comorbidities (C. C.)
60 Tonsillectomy and/or adenoidectomy only, age 0–17
89 Simple pneumonia + pleurisy age > 69 with C. C.
119 Vein ligation + stripping
127 Heart failure + shock
162 Inguinal + femoral hernia procedures age 18–69 w/o C. C.
167 Appendectomy w/o complicated principal diag age < 70 w/o C. C.
182 Esophagitis, gastroentestinal + miscellanous digestive disorders age > 69 with C. C.
183 Esophagitis, gastroentestinal + miscellanous digestive disorders age 18–69 w/o C. C.
184 Esophagitis, gastroentestinal + miscellanous digestive disorders age 0–17
209 Major joint + limb reattachment procedures
219 Lower extremity + humeral procedures/except hip, foot, femur age 18–69 w/o C. C.
222 Knee procedures age < 70 w/o C. C.
225 Foot procedures
231 Local excision + removal of int fix devices except hip + femur
243 Medical back problems
281 Trauma to the skin, subcut tiss + breast age 18–69 w/o C. C.
324 Urinary stones age < 70 w/o C. C.
355 Uterine, adnexa procedures for non-ovarian/adnexal malignancy age < 70 w/o C. C.
369 Menstrual + other female reproductive system disorders
371 Cesarean section w/o C. C.
373 Vaginal delivery w/o complicated diagnosis
381 Abortion with dilatation + curettage, aspiration curettage or hysterectomy

References

1. *Arons RR.* The new economics of health care: DRGs, case mix and patients length of stay. New York: Praeger, 1987.
2 *Blaco R.* Analisi esploratoria degli outliers. In: SAS User Group Italia, Atti del 7° convegno annuale, Firenze: SUGITALIA 91, 1991.
3 *Eggli Y, Grimm R, Paccaud F.* Adaptation du "Grouper" aux Statistiques médicales VESKA (1980–1986), (Cahier de Recherches Doc. IUMSP No 22). Lausanne: IUMSP, 1987.
4 *Gross AJ, Clark VA.* Survival distributions: reliability applications in the biomedical sciences. New York: Wiley, 1975.
5 *Marazzi A.* Algorithms, routines and S functions for robust statistics. Pacific Grove, Ca: Wadsworth & Brooks/Cole, 1992.
6 Commission on professional and hospital activities (CPHA). Length of stay in PAS hospitals by diagnosis, USA, 1977. Ann Arbor: CPHA, 1979.
7 *Berki SE, Ashcraft MLF, Newbrander SC.* Length of stay variations within ICDA-8 DRGs. Med Care 1984; *22*:126–142.
8 *Koehn V, Eggli Y, Paccaud F.* Comparaisons internationales des durées de séjour hospitalières. Méd et Hyg 1990; *48*:385–388.
9 *Cleary PD, Greenfields S, Mulley AG, Pauker SG, Schroeder SA, Wexler L, McNeil BJ.* Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. JAMA 1991; *266*:73–79.
10 *Mushlin AL, Black ER, Connolly CA, Buonaccorso KM, Eberly SW.* The necessary length of stay for chronic pulmonary disease. JAMA 1991; *266*:80–83.
11 *Epstein AM, Bogen J, Dreyer P, Thorpe KE.* Trends in length of stay and rates of readmission in Massachusetts: implications for monitoring quality of care. Inquiry 1991; *21*:19.
12 *List ND, Fronczack NE, Gottlieb SH, Baker RE.* A cross-national study of differences in length of stay of patients with cardiac diagnoses. Med care 1983; *21*:519–530.

Address for correspondence:
Christiane Ruffieux
Institut universitaire de medecine sociale et préventive
Rue du Bugnon 17
CH-1005 Lausanne/Suisse