Health Services Management Research **20:** 203–210

New approaches to reimbursement schemes based on patient classification systems and their comparison

Alfio Marazzi*, Lucien Gardiol† and Hong Dung Duong‡

**Institute for Social and Preventive Medicine; †Institute of Health Economics and Management, University of Lausanne; ‡Financial Office, University Hospital of Lausanne, Lausanne, Switzerland*

We propose reimbursement schemes based on patient classification systems (PCSs) that include adjustments for length of stay (LOS) and exceptional costs and are designed to minimize undesirable effects of economic incentives. In addition, a statistical approach to compare the schemes and the underlying PCSs is proposed, where costs and LOSs for two successive years are used. The first year data provides estimates of the class cost means and the next year's reimbursements which are compared with the second year's costs. This method focuses on the predictive power of a PCS and differs from the usual retrospective analyses based on the proportion of explained variance for single year data.

The approach is applied to discharge data of Swiss hospitals where stays are grouped according to five PCSs: All Patient Diagnosis-Related Groups (AP-DRGs), All Patient Refined Diagnosis-Related Groups (APR-DRGs), International Refined Diagnosis-Related Groups (IR-DRGs), Australian Refined Diagnosis-Related Groups (AR-DRGs), and SQLape.

When adjusting for LOS and outliers, these systems do not differ substantially in their ability to predict cost of stay. Therefore, increasing the number of classes does not necessarily improve cost predictions. However, the payment of a fixed amount per diem (not exceeding the marginal cost) and correcting the reimbursements for exceptional costs substantially reduces the average discrepancy between costs and reimbursements.

Introduction

Patient classification systems (PCSs) are an important tool of hospital management. A PCS is a set of rules that ascribes each individual hospital stay to a *class* or *group*. The hospital population is then a mixture of cases, i.e. a case-mix, which is characterized by the proportion of cases in each class. The rules for ascribing a case to a class are determined so that the classes are as homogeneous as possible with respect to clinical criteria (e.g. diagnoses and procedures) and to resource consumption. In routine applications, cases are ascribed to classes with the help of a computer program called a grouper. The classes are usually called diagnosis-related groups or DRGs.

PCSs are used for epidemiological monitoring and for standardized comparison of hospital activity. The main applications are however hospital budgeting, funding, and reimbursement. For this purpose, a mean cost of each

Alfio Marazzi PhD, Institute for Social and Preventive Medicine, University of Lausanne, Bugnon 17, CH 1005 Lausanne, Switzerland; **Lucien Gardol PhD**, Institute of Health Economics and Management, University of Lausanne, Switzerland; **Hong Dung Duong**, University Hospital of Lausanne, Switzerland
Correspondence to: Alfio Marazzi
E-mail: alfio.marazzi@chuv.ch

class is usually estimated with the help of available information on a national basis. The cost means are then used as a basis for negotiation to fix prospective payment or retrospective reimbursement rates.

The design and development of PCSs began in the late 60s at Yale University. The concept then migrated from the USA to Western European countries, Australia, Eastern Europe, and Asia.[1] This dissemination gave rise to a huge variety of PCSs with different class definitions and class numbers, ranging from a few hundreds to a few thousands.[2] Unfortunately, a general methodology to evaluate and compare PCSs is still lacking. This makes the choice of an appropriate grouper and reimbursement rule a very difficult problem for health managers.[3]

An additional problem is due to the presence of outliers in hospital data. Outliers are atypical cases usually due to specific complications not captured by the classification system and often associated with an extraordinarily long length of stay (LOS) and high costs. Unfortunately, common statistical measures can be greatly distorted by outlying values. Outlier resistant (robust) procedures are, therefore, required for the computation of basic DRG means[4–7] as well as for the assessment of classification schemes.[8] Moreover, special rules for determining payment for exceptional costs must be conceived.[9]

This paper proposes a simple statistical approach to compare classification systems and prospective reimbursement schemes. Firstly, reimbursement schemes that include adjustments for exceptional costs and LOS are proposed. The schemes are designed to minimize undesirable effects of economic incentives, such as providing insufficient care or premature patient dismissal. Both the reimbursement schemes and the statistical approach to compare them are based on cost and LOS for two successive years (or periods). The stays for the first year are classified according to a given classification system. Then, they are used to compute robust estimates of the class cost means and to define reimbursements for the next year. The discrepancies between the reimbursements and second year observed costs provide a basis for comparing systems and reimbursement schemes.

The approach is tested on a database of 127,838 stays (67,137 in 2002 and 60,701 in 2003) in a Swiss hospital collective. Four classification systems—All Patient Diagnosis-Related Groups (AP-DRGs, version 12), International Refined Diagnosis-Related Groups (IR-DRGs, version 1.2), All Patient Refined Diagnosis-Related Groups (APR-DRGs, version 15.0)—provided by 3M Health Information Systems,[2] and Australian Refined Diagnosis-Related Groups (AR-DRGs; version 5.1; www.health.gov.au/casemix). In addition, a new PCS based on the same hospital discharge data set used by DRGs and called SQLape,[10] has been tested. Contrary to DRGs that allow only one group for each stay, SQLape assigns stays to multiple 'categories' of diseases or operations. The number of categories is determined by about 180 medical and 180 surgical conditions. Approximately 75% of Swiss stays are assigned to a single SQLape category, 17% to two categories, 5% stays have three, and less than 3% more than three categories.

Methods

Retrospective versus prospective reimbursement systems

Using a retrospective system, the rates are determined at the end of the reimbursed period (e.g. a year) based on observed cost statistics. With a *basic prospective system*, class standardized rates are determined at the beginning of each period based on the *class cost means* of the preceding period. Retrospective systems are usually validated with the help of ordinary regression (OLS) of current costs against classes and the associated coefficient of determination R^2. We evaluate prospective systems by comparing second period individual costs to the appropriate class rates (a detailed description of the method follows).

Selection and incentive

According to Newhouse,[11] the main positive effect of a PCS-based reimbursement system is that cost reducing incentives are maximized, as the hospitals completely benefit from any cost reducing activity. In addition, payments are distributed fairly between hospitals, reimbursements are related to case severity, and insurers contribute according to the effective utilization of hospital services by those insured. Unfortunately, incongruent incentives may also be introduced. In particular, hospitals are tempted to select patients, refusing those that might be very expensive. We observe that gains from selection can only occur if selection is based on information not captured by the PCS and, in this case, cost varies in a predictable way within classes. Therefore, we focus on comparing intra-class variances: the lower the intra-class variance, the lower the selection opportunities.

Length of stay

A pathology-based reimbursement system should bind the incentive to keep patients longer than they need. However, if the system is completely LOS independent, it may induce premature dismissal. Therefore, we consider a modification of the basic reimbursement scheme that includes the payment of a fixed amount per diem. We keep this amount lower than the marginal cost per diem (i.e. the cost that would be generated by a healthy patient) in order to discourage early dismissal, without any sequelae of unnecessarily long inpatient stays.

Data quality and outliers

Outliers may have different origins, such as coding errors, bad classification, or other unexplained sources. Coding errors are independent of the classification systems and have, therefore, a minor impact on system comparisons. The main impact of outliers on the accuracy of classification systems is as a result of bad classifications.

In a prospective setting, we distinguish first and second period outliers. First period outliers may influence basic means and standardized rates. In order to limit the outlier effects on cost means, the proposed schemes use a particular kind of robust statistic called *truncated means* (see below for a technical definition). Second period outliers, i.e. exceptional costs to be reimbursed, will be called *exceptions*. We consider a second modification of the basic reimbursement system that includes a special reimbursement rule for exceptions. This rule is closely related to stop loss insurance, as in Ma.[12] A simple stop loss rule would reimburse the complete cost exceeding a certain amount inducing hospitals to maximize costs, whereas the proposed rule pays only a fraction of the cost that exceeds the truncated mean by more than a given threshold.

In practice, the threshold has to be defined on the basis of the first period cost distribution. However, in order to compare classification systems, another definition is used in this paper. Since special reimbursements for exceptions obviously decrease the reimbursement errors and favour bad classification systems (with many errors), the threshold is defined so that the number of outliers is the same (e.g. 1000) for all systems (a formal definition follows).

Inflation

Health-care costs have a propensity to rise. In practice, econometric tools are used to detemine the appropriate factor to inflate first period means and obtain second period rates. Since we compare grouper accuracy—not their ability to predict inflation—we use inflation factors (retrospectively computed) for which the costs of the second period are completely reimbursed.

The truncated mean (technical definition)

The truncated mean is a robust estimate of the mean of an asymmetrically distributed random variable. It is based on a simple outlier rejection rule and can be used to estimate a mean cost of stay in the presence of atypical cases (too expensive or too cheap).

Let $x_1, \ldots, x_n$ be a sample of the cost of stay, for a certain group of stays. We assume that the distribution of the random variable cost of stay x can be described by a parametric model (e.g. a gamma or a lognormal density) with parameters τ and σ, density $g_{\tau,\sigma}$, and cumulative distribution function $G_{\tau,\sigma}$. Such models provide adequate descriptions of most DRGs.[6] Let t and s be the empirical median and median absolute deviation. The model median and median absolute deviation can be expressed with the help of two functions med(τ,σ) and mad(τ,σ) of the parameters. Thus, τ and σ can be determined by solving the two equations:

$$\mathrm{med}(t,\sigma) = t \text{ and } \mathrm{mad}(t,\sigma) = s. \tag{1}$$

Let u be a user chosen number close to 1, e.g. $u = 0.99$ and T_u be the u-quantile of $G_{\tau,\sigma}$. We truncate $g_{\tau,\sigma}$ at T_u (*upper truncation point*) and T_l (*lower truncation point*) and choose T_l so that the mean of the truncated model equals the mean of the entire model, i.e.

$$\int_{T_l}^{T_u} x g_{\tau,\sigma}(x)\,\mathrm{d}x / (u - G_{\tau,\sigma}(T_l)) = \int x g_{\tau,\sigma}(x)\,\mathrm{d}x. \tag{2}$$

The *truncated mean* estimate is then defined as

$$m(u) = \text{mean}\{x_i \text{ such that } T_l < x_i \leq T_u\}. \quad (3)$$

Clearly $m(1)$ is the usual arithmetic mean. Equations (1) and (2) can be solved with the help of simple computer programs; particular implementations are provided by the authors.[13] In this paper, we used a gamma distribution model for all groups.

The truncated mean is used to estimate cost means for the AP-, APR-, IR-, and AR-DRG systems. SQLape has its own way of estimating the expected costs per hospitalization. It uses a sum of points per operation or medical condition, which is reduced if the patient is transferred to another hospital or dies before the expected discharge date. Attention is paid to minimize differences in coding practice between providers. In particular, the main categories do not depend on the principal diagnosis selected by the physician in order to hinder 'DRG creeping'.[10]

Formal definition of the reimbursement rule

We suppose that LOSs and costs are available for two successive years, say 0 and 1. Let x^0_{gi} ($g=1, \ldots, G$, $i=1,\ldots, n^0_g$) denote the cost of stay i in group g and year 0 and x^1_{gj} ($g=1, \ldots, G$, $j=1,\ldots, n^1_g$) the cost of stay j in group g and year 1. Similarly, y^0_{gi} and y^1_{gj} denote LOSs. The reimbursement of x^1_{gj} is determined on the grounds of year 0 observed costs and LOSs of group g according to the following general scheme.

We assume that the reimbursement includes a fixed daily amount d; thus, the reimbursement to stay j in group g and year 0 includes an LOS proportional component dy^0_{gj}. Let

$$m^0_g(u,d) = \text{mean}\{x^0_{gi} - dy^0_{gi} \text{ such that } T_{lg} < x^0_{gi} - dy^0_{gi} \leq T_{ug}\}$$

be the LOS adjusted truncated cost mean of class g in year 0. Here, T_{ug} and T_{lg} are determined according to equations (1)-(2) using the adjusted costs $x^0_{gi} - dy^0_{gi}$ of class g. A typical stay in class g and year 1 receives an amount equal to $m^0_g(u,d)$ plus a sum proportional to its LOS. In addition, an extra amount for exceptional costs is added if the actual adjusted cost $x^1_{gj} - dy^1_{gj}$ exceeds $m^0_g(u,d)$ by more than a given threshold h. On the other hand, the reimbursement is reduced if the actual adjusted cost is lower than $m^0_g(u,d) - h$. Note that the single threshold h is not directly related to the group truncation limits T_{ug} and T_{lg}. Therefore, an exceptional cost exceeding $m^0_g(u,d) + h$ is not necessarily rejected in the computation of $m^0_g(u,d)$, and conversely. For this reason, we call *outlier* any stay such that $x^0_{gi} - dy^0_{gi} \notin [T_{lg}, T_{ug}]$ and *exception* any stay such that $x^1_{gj} - dy^1_{gj} \notin [m^0_g(u,d) - h, m^0_g(u,d) + h]$.

More precisely, let u, h, k, d, and f be parameters chosen by the user (see below). The *general reimbursement* of stay j in class g is given by $r^0_{gj} = r^0(x^1_{gj}, y^1_{gj}; u, h, k, d, f, g)$ where the function r^0 is defined by

$$r^0(x, y; u, h, k, d, f, g) = f[r(x - dy, m^0_g(u, d), h, k) + dy],$$

and the function r is given by (Figure 1)

$$\begin{aligned} r(x, m, h, k) & \\ &= m && \text{if} \quad m - h \leq x \leq m + h, \\ &= m + k(x - m - h) && \text{if} \quad m + h < x, \\ &= m - k(m - h - x) && \text{if} \quad x < m - h. \end{aligned}$$

Special cases

Setting $d=0$, we obtain a reimbursement scheme independent of LOS. For $k=0$, no correction for exceptions is made. Thus, for $d=k=0$, all stays in DRG g receive the same reimbursement $fm^0_g(u,0)$. For $u=1$, the arithmetic mean is used in place of the truncated

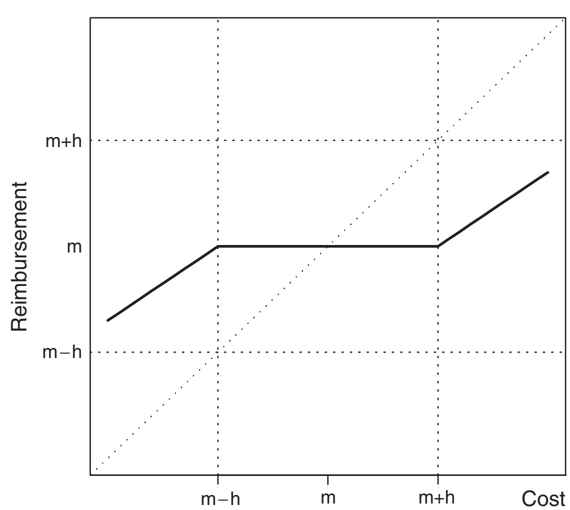

Figure 1 *Reimbursement function. The broken line represents the function r = x (r = reimbursement; x = cost). The full line represents the function r(x,m,h,k) for given m = 10, h = 5, and k = 0.9*

mean; in particular,

$$r^0(x^1_{gj}, y^1_{gj}; 1, h, 0, 0, f, g) = f\, mean\{x^0_{gi}\}.$$

Tuning

The parameters u, h, k, d, and f must be chosen by the user. The *upper probability of non-rejection* u is usually determined according to statistical criteria; $u=0.99$ is very common. The *cut-off* h, as well as the *damping factor* k for exceptions, are arbitrary; in general, k is positive and close to 1 (e.g. 0.95) and h is chosen so that the number of exceptions is not too high (e.g. 1 per 1000 stays, so that a thorough outlier inspection is made possible). The *fixed reimbursement* d must not exceed the marginal cost per diem (e.g. 200 Swiss francs CHF). The constant f is interpreted as an *inflation factor* and should be determined with the help of standard econometric tools.

Measures of reimbursement performance

In general, the errors $e_{gj}=x^1_{gj}-r^0_{gj}$ provide information about the quality of the reimbursement scheme, and different classification-reimbursement schemes can be compared by analysing their *error distributions*. For example, when $k=0$, a long upper tail of the error distribution suggests a high frequency of significant misclassifications.

A simple measure of reimbursement performance is the *reduction in mean squared reimbursement error* defined as

$$R^2_r = 1 - \mathrm{M2R}/\mathrm{mean}\{(x^1_{gj} - \bar{x}^1)^2\},$$

where $\mathrm{M2R} = \mathrm{mean}\{e^2_{gj}\}$ is the *mean squared reimbursement error* and $\bar{x}^1$is the overall cost mean for year 1. This measure is very sensitive to extreme errors : large errors are penalized more than small errors. Therefore, we also use another measure called the *reduction in mean absolute reimbursement error*

$$R^A_r = 1 - \mathrm{MAR}/\mathrm{mean}\{|x^1_{gj} - \bar{x}^1|\}$$

where $\mathrm{MAR} = \mathrm{mean}\{|e_{gj}|\}$is the *mean absolute reimbursement error*. MAR is itself an interesting measure on the same scale as cost. R^A_r is less sensitive than R^2_r to extreme errors. A good classification system provides high values of R^2_r and R^A_r.

Remark

In the special case $d=k=0$ and $f=1$, we obtain:

$$R^2_r = 1 - \sum_{g,j}(x^1_{gj} - m^0_g(u))^2 / \sum_{g,j}(x^1_{gj} - \bar{x}^1)^2$$

and it can be shown that

$$R^2_r = R^2 - \sum_{g} n^1_g(m^0_g(u) - \bar{x}^1_g)^2 / \sum_{g,j}(x^1_{gj} - \bar{x}^1)^2 < R^2$$

where

$$R^2 = 1 - \sum_{g,j}(x^1_{gj} - \bar{x}^1_g)^2 / \sum_{g,j}(x^1_{gj} - \bar{x}^1)^2$$

is the usual proportion of explained variance and

$$\bar{x}^1_g = \mathrm{mean}_j\{x^1_{jg}\}.$$

Thus, R^2_r is the usual R^2 for year 1 data penalized for the last year 'bad' prediction, i.e. for the discrepancy between the class mean $\bar{x}^1_g$ and its prediction $m^0_g(u)$.

Data and results

We used a database of 67,137 stays in 2002 and 60,701 stays in 2003 of a set of Swiss hospitals in order to compare the classification systems AP-DRG, IR-DRG, APR-DRG, AR-DRG, and SQLape. For the year 2003 we found 623 AP-DRGs, 747 IR-DRGs, 1200 APR-DRGs, and 634 AR-DRGs. For each DRG system, a few groups were present in the year 2003 but not in the year 2002, and their mean cost was estimated using the available stays of the closest group or the corresponding major diagnostic category.

We considered the special cases of the general reimbursement scheme summarized in Table 1. The cut-off h and the factor f were determined for each system and each combination

Table 1 *Special cases of the general reimbursement scheme*

Reimbursement type	u	k	d	Adjustment
B (basic)	0.99	0.00	0	None
L	0.99	0.00	200	For LOS
E	0.99	0.95	0	For exceptions
L and E	0.99	0.95	200	For LOS and exceptions

of u, k, and d so that:

(a) Number of elements in

$$\{j \text{ such that } x^1_{gj} - dy^1_{gj} \notin [m^0_g(u,d) - h, m^0_g(u,d) + h]\} = 1000,$$

(b)

$$\sum_{g,j} r^0(x^1_{gj}, y^1_{gj}; u, h, k, d, f, g) = \sum_{g,j} x^1_{gj}.$$

Condition (a) ensures that the number of exceptions was 1000 and condition (b) that the totality of the costs was reimbursed. These conditions standardized comparisons so that they depended neither on the number of exceptions nor on the proportion of reimbursed total costs.

From the results collected in Table 2 we observe that:

(1) The (retrospective) R^2s are close to published values for the same DRG systems:[2] APR-DRGs have the highest R^2, IR-DRGs and AR-DRGs the lowest. This is no surprise, as R^2 increases with the number of groups. In addition, IR-DRGs, APR-DRGs, and SQLape seem to better capture the cost increase as their inflation factor is generally lower than for AP-DRGs. However, the reimbursement quality of AP-DRGs, IR-DRGs, APR-DRGs, and AR-DRGs – as measured by R^2_r, R^A_r and MAR – is virtually the same (by reimbursement quality we mean the ability to reduce the difference between the observed cost of a stay and the amount paid for this same stay). These results are not contradictory. As noted in the remark at the end of the previous section, the reimbursement quality is penalized by the imperfect predictions of the group cost means, and the stability of these predictions decreases with increasing group number (since fewer stays are observed in each group). We note that the number of stays of our collective of Swiss hospitals is low with respect to American standards, but not atypical for European countries. The basic reimbursement scheme based on SQLape performs better than the other systems; this advantage decreases when adjustments for LOS and exceptions are introduced.

(2) Both the adjustments for exceptional costs and LOS provide substantial gains in reimbursement quality. The most important increase in R^2_r is as a result of the

Table 2 *Results*

		R^2	$R_r{}^2$	R^A_r	MAR	f	h
AP-DRG	B	0.49	0.38	0.37	5112	1.225	—
	L		0.52	0.48	4258	1.203	—
	E		0.83	0.48	4215	1.158	36,200
	L and E		0.88	0.57	3469	1.143	30,905
IR-DRG	B	0.48	0.44	0.38	5095	1.191	—
	L		0.56	0.48	4262	1.171	—
	E		0.83	0.47	4302	1.132	35,080
	L and E		0.88	0.56	3563	1.118	29,715
APR-DRG	B	0.61	0.39	0.40	4922	1.171	—
	L		0.52	0.49	4133	1.153	—
	E		0.83	0.49	4137	1.126	34,965
	L and E		0.88	0.58	3433	1.113	29,460
AR-DRG	B	0.48	0.41	0.37	5103	1.208	
	L		0.54	0.48	4273	1.183	
	E		0.82	0.48	4269	1.149	36,700
	L and E		0.87	0.57	3543	1.132	31,470
SQLape	B	—	0.55	0.42	4726	1.031	
	L		0.64	0.54	3723	0.888	
	E		0.84	0.49	4191	1.008	31,810
	L and E		0.88	0.59	3338	0.883	27,272

adjustment for exceptional costs (because R_r^2 is more sensitive to extreme values than R_r^A); the gain in R_r^A, as a result of the two single adjustments, are almost equivalent. Their combined effect is larger than both their specific effects and corresponds to a reduction of about 1500 CHF in mean absolute error with respect to the basic reimbursement rule.

(3) For the AP-DRGs, IR-DRGs, APR-DRGs, and the AR-DRGs systems, the threshold h is about 35,000 CHF when the simple adjustment for exceptional costs is used and about 30,000 CHF when both the adjustments are used. This difference is explained by the fact that h was determined so that the number of exceptions was 1000 in both cases. With a unique threshold, the combined adjustment would clearly provide a lower number of exceptional cases. For a similar reason, the lower threshold values required by SQLape (32,000 and 27,000 CHF) clearly suggest that this system better captures exceptional cases. The values of h are very high, and therefore for 98% ($\approx$59,000/60,000) of the cases, the basic reimbursement rule is not affected by the correction for exceptional cases. However, the introduction of the special rule for exceptions would lower the amounts reimbursed to the typical cases by 0.5% (SQLape) to 5% (AP-DRG) in order to cover the exceptional costs.

Discussion

The statistical approach to compare reimbursement schemes introduced in this paper uses new measures of predictive power. These measures compare mean cost predictions, i.e. reimbursements based on a given period to the observed costs of a different period. They are the reduction in mean squared reimbursement error (R_r^2) and the reduction in mean absolute reimbursement error (R_r^A).

The traditional measure is the usual coefficient of determination R^2, which is based on the retrospective comparison of the data for a single period, with their own group means. Both R^2 and R_r^2 use the square penalty function, where large prediction errors are penalized much more than small errors. Both are very sensitive to extreme errors and one may be tempted to remove them. However, the purpose of these measures is not robust estimation of some population parameter, but the assessment of the reimbursement quality; excluding very expensive cases would, therefore, provide an over-optimistic indication. The absolute measure R_r^A takes all errors into account but penalizes them according to their absolute size. This seems to fit well with the payment context.

The differences between R^2, R_r^2, and R_r^A partly explain the results. For example, using R^2, the more complex APR system does well relative to the other systems. This system seems better at classifying less common but very expensive cases. However, both R_r^2 and R_r^A indicate that a system with a large number of classes does not necessarily provide a better cost prediction with respect to a simpler system, unless the number of stays in each class is large enough to provide a satisfactory statistical stability of the prediction.

Two features of a reimbursement scheme based on PCS, standardized rates provided significant benefits in reimbursement quality: (a) the payment of a fixed amount per diem, (b) the partial payment for any cost that exceeds the appropriate robust class mean by more than a given threshold. These features can be fine-tuned in order to lower the risks of undesirable economic incentives such as cost maximization or premature dismissal. Adjusting the payment method for exceptional costs and LOS greatly reduces the effects of prediction errors and, therefore, the impact of the classification system. We note that this significant improvement is obtained by changing the basic reimbursement rule of a very low percentage of the cases.

We may leave the conclusions to the user. If he or she feels (in the user's context) that a few large errors are much worse than many small errors, both the traditional R^2 and the reduction in mean squared reimbursement error may provide useful measures for comparisons. If, instead, the user feels that a few large errors are not worse than many small errors (with the same sum in terms of profit and losses), then the reduction in mean absolute reimbursement error will more closely reflect this preference. In either case, differences between classification systems can be reduced through the use of appropriate adjustment policies. In particular, adjustments for exceptional costs and LOS are very effective. Clearly, in selecting a classification system, a variety of factors – such as the ability to facilitate comparison with other areas, the ability to easily localize cost, the degree of control over future development, sensitivity to coding practices – have to be considered. However, insufficient statistical stability, due to the scarcity of data, may reduce the desirable benefits of a more complex system. In this case, and for the purpose of prediction, a moderate number of classes may become a very important factor.

Acknowledgements

We are grateful to Alex Randriamiharisoa and Danielle Meylan for the programming work and to Yves Eggli, Hervé Guillain, Jean-Claude Rey, and James C Vertrees for their insightful comments, suggestions, and coordination support. We thank 3M Health Information Systems, SQLape s.à.r.l., and the Department of Health and Aging of the Commonwealth of Australia for making available their classification systems within the framework of the SwissDRG project (www.swissdrg.org).

References

1 Roger France FH. Case mix use in 25 countries: a migration success but international comparison failure. *Int J Med Inf* 2003;**70**:215–19

2 Averill RF, Muldoon MS, Vertrees JC *et al*. The evolution of casemix measurement using diagnosis related groups (DRGs). 3M HIS Working Paper 5–98 1998

3 Palmer G, Reid B. Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. *Health Serv Manage Res* 2001; **14**:71–81

4 Cots F, Elvira D, Castells X, Saez M. Relevance of outlier cases in case mix systems and evaluation of trimming methods. *Health Care Manag Sci* 2003;**6**:27–35

5 Marazzi A, Ruffieux C. Rules for removing outliers and tests based on parametric models for estimating and comparing average length of stay. *Patient Classification Systems/Europe (PCS/E), 14th International Working Conference Proceedings Manual*. Manchester, 1998: 191–200

6 Marazzi A, Paccaud F, Ruffieux C, Beguin C. Fitting the distributions of length of stay by parametric models. *Med Care* 1998;**36**:915–27

7 Marazzi A, Ruffieux C. The truncated mean of an asymmetric distribution. *Comput Stat Data Anal* 1999; **32**:79–100

8 Kulinskaya E, Zhang X. A robust method for reviewing variance explained by diagnosis related groups. *Patient Classification Systems/Europe (PCS/E), 14th International Working Conference Proceedings Manual*. Manchester, 1998:185–90

9 Centre for Medicare & Medicaid Services. Medicare program; change in methodology or determining payment for extraordinarily high-cost (cost outliers) under acute care hospital inpatient and long-term care hospital prospective payment systems. Final rule. *Fed Regist* 2003;**68**:34493–15

10 Eggli Y. SQLape®: Prévision des coûts hospitaliers fondée sur le profil des patients, SQLape sàrl, 2005. Available at www.sqlape.com

11 Newhouse JP. Reimbursing health plans and health providers; efficiency in production versus selection. *J Econ Lit* 1996;**34**:1186–263

12 Ma CA. Health care payment systems: cost and quality incentives. *J Econ Manage Strategy* 1994;**3**:93–112

13 Marazzi A, Randriamiharisoa A, Ruffieux C. Splus functions for truncated and weighted means of asymmetric distributions: Weibull, Gamma, and Lognormal cases. *Technical report, Institut Universitaire de Médecine Sociale et Préventive*. Lausanne, 1999. Available at www.iumsp.ch