A Multi-level System of Performance Evaluation Using Diagnosis-Related Groups for Cost Containment

Shuguang Lin[1,2] | Paul Rouse[2,*] | Fan Zhang[3] | Ying-Ming Wang[1,4,*]

[1] Decision Sciences Institute, Fuzhou University, Fuzhou 350108, China

[2] Department of Accounting and Finance, University of Auckland, Auckland 1142, New Zealand

[3] Fujian Medical University Affiliated Fuzhou First Hospital, Fuzhou 350009, China

[4] Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350108, China

*Corresponding author:

Ying-Ming Wang, Decision Sciences Institute, Fuzhou University, 2 Xueyuan Road, Fuzhou 350108, China, email: ymwang@fzu.edu.cn, Tel: +86 591 228 66681.
Paul Rouse, Department of Accounting and Finance, University of Auckland, Owen G Glenn Building, 12 Grafton Road, Private Bag 92019, Auckland 1142, New Zealand, email: p.rouse@auckland.ac.nz, Tel: +64 9 923 7192, Fax: +64 9 373 7406

Abstract

This study aims to develop a performance evaluation system that can facilitate performance evaluation at region, hospital, and department levels to enable better cost management for sustainable development. A multi-level system of performance evaluation informs a hierarchical assessment of cost management from regions to hospitals to departments using diagnosis-related group (DRGs). Various metrics are developed employing the variances between targets and actuals where targets are determined from two perspectives: benchmarking using external regional prices and change management using internal data. Targets for the latter are statistically based and specifically incorporate variability. The model is applied to two hospitals, twenty departments, nine DRGs and 1071 inpatients. The analyses indicate that the approach can provide a practical evaluation tool that allows for particular characteristics at multiple levels. The system provides macro-micro and external-internal perspectives in performance, enabling high-level variances to be decomposed thereby identifying sources of performance variability and financial impact.

Keywords: Cost containment, Performance Evaluation, Multi-level System, Diagnosis-related Group (DRG), Health system sustainability

1. Introduction

Traditionally, countries have used two basic mechanisms to pay for hospital care: fee-for-service payments and global budgets. Both are problematic in terms of ensuring high-quality care due to the inherent incentive to overprovide (fee-for-service) or under-provide (global budgets) hospital services [1]. In recent years, many countries have introduced policies around fixed price payments for each patient, prospective reimbursement systems, which are considered to be more cost-efficient than cost-based retrospective reimbursements in response to rapid growth in the cost of healthcare. Diagnosis-related groups (DRGs) are a key component of these systems intended to control costs and motivate healthcare providers to improve efficiency [1]. DRGs provide multi-product 'views in acute care hospitals, classifying patients with similar resource consumption patterns into clinically meaningful categories [2]. DRGs can be used for hospital performance comparisons and to promote hospital management, as DRG-based hospital payment systems [3]. These payment systems have the objective of incentivising hospitals to improve their performance [4].

In contrast to cost-reimbursement, however, DRG prospective payment systems (DRG-PPS) decouple hospital revenue from a largely fixed, hospital cost base, thereby transferring significant financial risks from the system level to the hospital level. Hospitals are forced out of an almost risk-free cost reimbursement system and face uncertain financial consequences for their health care provision. The U.S. healthcare sector has shifted toward value-based payment models, and hospitals face increasing pressure to reduce costs and improve the efficiency of care delivery [5]. In German, the introduction of a performance pay component with bonus–malus incentives to the DRG system has a positive effect on the quality of care [6]. China commits to achieve Healthy China 2030, a statement of political will to prioritise population health and respond to global commitments related to realising the United Nations sustainable development goals [7]. In South Korea, the introduction of a new DRG-based payment system using a conventional top-down approach has created a large perception gap between doctors and health policy makers [8].

The ultimate goal of health reform is to achieve sustainable cost containment and improve the performance of the healthcare system [9]. In order for a prospective payment policy to succeed, how to objectively and equitably measure the cost containment performance of healthcare providers is important from both cost efficiency and effectiveness perspectives. Hospital managers need to not only monitor cost containment performance for each DRG and clinical department to ensure that they can operate under a PPS but also consider incentives for departments and medical staffs to improve cost management. This paper describes a multi-level model that uses external reference prices for benchmarking that link to internally generated targets for change management. This can be envisaged in two parts where first hospitals need to ensure their costs are in line with other hospitals and sector prices. Second, hospital management needs to monitor internal costs at greater levels of granularity and identify targets that will motivate and change employee behaviour where necessary while maintaining quality standards.

The paper contributes as follows. First, a multi-level evaluation system is described to enable high-level variances to be decomposed into hospital and departmental variances that encompass both an external funding benchmarking perspective and an internal, change behaviour objectives. These inform the identification and estimation of appropriate targets for hospital senior management and at departmental levels. Second, while a PPS policy was the motivator for this system, it can also be used for any regime where external DRG prices are available for benchmarking purposes. Third, DRG specific metrics of performance are developed to quantitatively assess the performance comprising total, mean, percentage and

normalised variances. These are illustrated using two hospitals, nine DRGs, twenty departments and 1071 inpatients to demonstrate the principles, advantages and disadvantages of the system. Fourth, the trimmed lognormal mean and standard deviation in light of the probability distribution of health cost are used to set the cost targets for internal purposes as part of efforts to change behaviour. For some departments such as intensive care unit (ICU), low volumes for some DRGs may require a different approach and we suggest some ways to do this. The results reveal how costs can vary depending on the departmental structure and patient complexity with high cost mostly caused by ICU including PICU (Paediatric Intensive Care Unit).

The paper is structured as follows: Section 2 reviews the development of DRG-based performance evaluation including case-mix, DRG costing and cost management; Section 3 introduces the methodology, a multi-level system and the metrics for measuring performance; Section 4 illustrates the system with performance evaluations of two hospitals; Section 5 discusses the utility of performance evaluation based on DRGs, and a research agenda to assess to address some crucial problems in hospital performance evaluation; Section 6 concludes this study and points out limitations and future research.

2.   Literature reviews
2.1 Performance evaluation with DRGs

DRGs have been used for many purposes including prospective payment systems, risk adjustment and performance evaluation of work efficiency and medical service quality [10]. Park and Shin [11] developed a model that measures and evaluates the case-mix complexity of tertiary care hospitals. DRGs are used to evaluate the inpatient service performance of the clinical subspecialty "major operation of the digestive system" of a cancer specialist hospital [12]. The relationships among the average number of procedures per surgeon, across the entire hospital, and multiple dimensions of overall hospital operational performance are evaluated using All Patient Refined Diagnostic Related Groups (APR-DRGs) to adjust for case-mix and severity [13]. DRGs are still predominantly used in most studies on hospital performance and case-mix complexity. Actual improvements in cost per discharge[1], average length of stay, and conformance quality are used to assess changes in a hospital's level of process excellence [14].

2.2 DRG costs

High-quality cost accounting systems were conducive to the development of DRG payment mechanisms [15]. The accuracy of the calculation is only one of the factors that determine the choice of the cost accounting methodology, important aspects include the selection of reference hospitals, individual country regulations and the complexity of hospital information systems in hospitals [16]. Internationally there are differences in cost accounting systems through cross-country comparisons to identify and improve areas of weakness [17]. The implementation of activity-based costing (ABC) was applied at a Heart Centre in Sweden for administrative cost information, strategic decision-making, quality improvement, and cost reduction [18].

There is also a tension between bottom-up and top-down costing where the former may provide different costs than the latter. Bottom-up microcosting, top-down microcosting and gross costing compared to estimate the total cost for appendectomy, normal delivery, stroke and Acute Myocardial Infarction (AMI) of hospital service that restricting the use of bottom-up microcosting to cost components with a large impact on total costs (i.e., labour and inpatient stay) would likely result in reliable cost estimates [19].

DRG costs are used to derive case weights as a basis for hospital funding and comparative performance evaluations. The consequences of DRG cost weight volatility on hospital

---

[1] Similar to DRGs but calculated differently in their study.

performance using the hospital base rate as a quantitative measure of hospital performance are estimated [20]. The average case weight increases by 14% for every one-day increase in the average length of stay (ALOS) [21]. The German DRG-system revealed that a "compression'' of DRG cost weights occurs and that the data sample used to calculate case weights lacks representativeness [22]. A better classification system and funding formulas incorporating reliable case weights derived from patient costing should overcome many of the deficiencies in the current case-mix payments systems [23].

2.3 Managing DRG-based costs

If a hospital could control the length of stay (LOS) at the projected level, on average, the cost per admission and the cost per patient day would decrease [24]. A significant reduction of LOS in hospitals that changed from per diem to DRG reimbursement in Switzerland [25]. Mean hospital cost as a function of patient stay is estimated and adjusted for the influence of patient characteristics and treatment procedures on LOS and cost [26]. The median cost for liver transplantation was not statistically significant with the corresponding hospital refund, the average DRG-based cost of liver transplantation was significantly higher than the actual cost, and the hospital's reimbursement for liver transplantation did not differ significantly from the actual registered cost [27]. From the hospital administration's perspective, it is not the additional costs of healthcare, but rather the cost-reimbursement relationship which guides decisions [28].
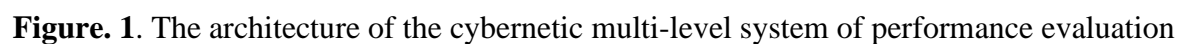
In summary, DRGs are commonly used for prospective payments systems and performance evaluation. While cost systems may differ for the determination of DRG costs and consequently case weights, there is likely to be reasonable uniformity within a specific country's health systems. Hence they form reasonable reference points for benchmarking purposes. So far as change behaviour is concerned, there appear to be established relationships between costs, patient stay and characteristics and conformance quality. The suggested use of an inter-quartile range to identify outliers is particularly relevant when considering cost containment especially for DRGs with high cost variability.

3.   Methodology

3.1 Architecture of the multi-level system

In benchmarking, a target is set based on other organisations in a relevant sector which is then used to compare with actual costs and possibly value engineer the product to match with the target. In this study, the regional target sets a reference for each hospital which is cascaded down to hospitals to departments through a top down process, such that the competitive pressure faced by hospitals can be traced to departments. Thus senior hospital management can identify how the hospital is performing at all levels against the regional benchmark (RB) as shown on the left hand side of the multi-level performance evaluation system depicted in Figure 1.

Target costing also considers prices but its focus is not so much to minimize cost, but to achieve a desired level of cost management determined by the target costing process. In healthcare this process must also maintain or improve quality. For change management (CM), the target is determined based on internal measures shown on the right hand side of Figure 1. This is a bottom up process where targets or prices are set for DRGs taking into consideration each department's and hospital's context. For example, ICU departments may have a different target setting method than general departments, tertiary hospitals are different from secondary hospitals.

Figure. 1. The architecture of the cybernetic multi-level system of performance evaluation

These department targets are then rolled upwards into hospital targets and thence upwards further to the region total targets . Variances can be obtained for both regional benchmark and change management with actual costs being the common component for both. At the hospital level three variances are significant. First the hospital benchmark variance shows whether the hospital is making or losing money against regional prices. Second the change management variance shows how internal target costs compare with actual in terms of cost management efforts. The third variance closes the loop by showing how far off internal target costs are from external reference prices. If a hospital is serious about change management one would expect the change management target cost to be less than actual costs. The size and sign of the third variance show the level of difficulty for a hospital relative to external pricing. In an ideal situation, this would be a favourable variance indicating that the hospital is performing well in its change management. An unfavourable variance indicates serious issues for future funding. The regional benchmark discerns where there is a profit, loss or break-even, the change management evaluates the internal performance considering their different structures and gives the directions to optimize the health system.

3.1.1 Evaluation for benchmarking

The revenues of hospitals under DRG payment schemes primarily derive from reimbursement by healthcare insurers which are determined by the DRG prices (or reimbursement rates) and volumes. DRGs prices are often based on the regional/national average DRG cost [22]. In Netherland, Poland and New Zealand, they are determined nationally while in China, it is done by region.

Thus, if the regional average cost is an appropriate target, it should cascade down to hospitals and departments. The cost target C_{jH}^{T} and standard deviation target σ_{jH}^{T} at the hospital level, and the cost target C_{jD}^{T} and standard deviation target σ_{jD}^{T} at the department level should be:

$$\mu_{j_R} = C_{j_H}^{\mathrm{T}} = C_{j_D}^{\mathrm{T}} \tag{1}$$

$$\sigma_{j_R} = \sigma_{j_H}^{T} = \sigma_{j_D}^{T} \tag{2}$$

where μ_{j_R} denotes the average cost of Region / Country R and σ_{j_R} is the standard deviation of the Region / Country. These targets can be compared against internal costs to reveal areas of strength and weakness across all levels.

3.1.2 Evaluation for change management

For benchmarking, we take the regional mean as the target, but this does not consider the circumstances of hospitals such as tertiary or secondary or departments that are more complex than others e.g. ICU. If the costs of some DRGs in a hospital are inherently lower than the regional mean (target) due to its different positioning, speciality, history or other causes, then the target may not be appropriate or even a disincentive to improve performance. Alternatively, the costs of some DRGs may be higher than the target if the hospital or department receives more complex patients.

Emmanuel et al. [29] make the following observation based on Hofstede (1968)'s study: "the budget level which motivates the best performance is one that is somewhat more demanding than the level of performance that will actually be achieved. However, a budget that is likely to be achieved will motivate only a lower level of performance". One of the major problems in hospital management is variability manifested in LOS, theatre times, delays. While some of this is due to the diversity of patient needs, considerable efforts are made to standardise processes where possible to decrease heterogeneity. Examples of these include case-mix systems, care bundles, specialist departments and standard operating procedures. Based on this, we argue that an internal performance evaluation system should incorporate variability as one of its components. The objective is to motivate departmental managers to explore better ways of managing patient flow and clinical activities.

Therefore one possibility is to set the target using some type of average cost (e.g. mean, median) for a DRG within the hospital or department adjusting for a proportion (θ) of the standard deviation (SD): Cost Target = Average Cost $- \theta \times$ SD. For target setting it is important that the target is perceived as fair and appropriate so we propose that the DRG costs are trimmed for both high and low outliers.

This has several advantages. First, the lower the standard deviation the closer the target becomes to actual cost thereby motivating departmental managers to find ways to reduce variability. Second, trimming by removing high and low outliers should be perceived as a fairer target than using the actual mean cost which includes these. Third, many researchers have noted that health care costs are skewed and not normally distributed [30-32]. Thus patient cost distributions can be heavily skewed to the right, signifying a form of cost distribution in which the mean value could lead to distorted targets. As several authors have pointed out, the cost function distribution is lognormal suggesting that the target should be based on a lognormal distribution [30,32,33]. Finally it should be noted that the comparator for both benchmarking and change management are actual costs. This ensures that managers are aware that these need to be managed and controlled as well as being familiar to them.

The next issue then arises as to the determination of θ and the percentage for trimming. A characteristic of a lognormal distribution is that when the logarithms of values form a normal distribution, the original values have a lognormal distribution. If the mean of the logarithms is set as the average and the adjustment factor θ is set at zero then departments will achieve the target (50% of the time); if θ is set at 0.1 then it will be less than 50% of the time and so on. The setting of θ is more dependent on management objectives, which will be demonstrated in Section 4 and further discussed in Section 5.

Spanish and Belgian studies have shown that high deficit cost outliers account for roughly

five percent of cases but produce 11–20 percent of inpatient costs [32,34]. Other studies suggest that nearly five percent of cases fall in a high outlier category except for very low-volume groups, with a recommendation that the percentage should not exceed ten percent [30,35]. Therefore in our example we trim the top five percent and the lowest five percent of cases.

A final issue concerns the volume of cases at departmental levels. For example, our empirical study was for only one month so that the number of cases for some DRGs was low even although on an annual basis, they would be much higher. Notwithstanding there are some complex DRGs that are low volume even on an annual basis; these tend to be cases in ICUs. Figure 2 sets out our thoughts on managing targets for low volumes and for specialized departments bearing in mind that the target setting process is likely to be done annually so that the incidence of low volumes for some DRGs would likely be low.
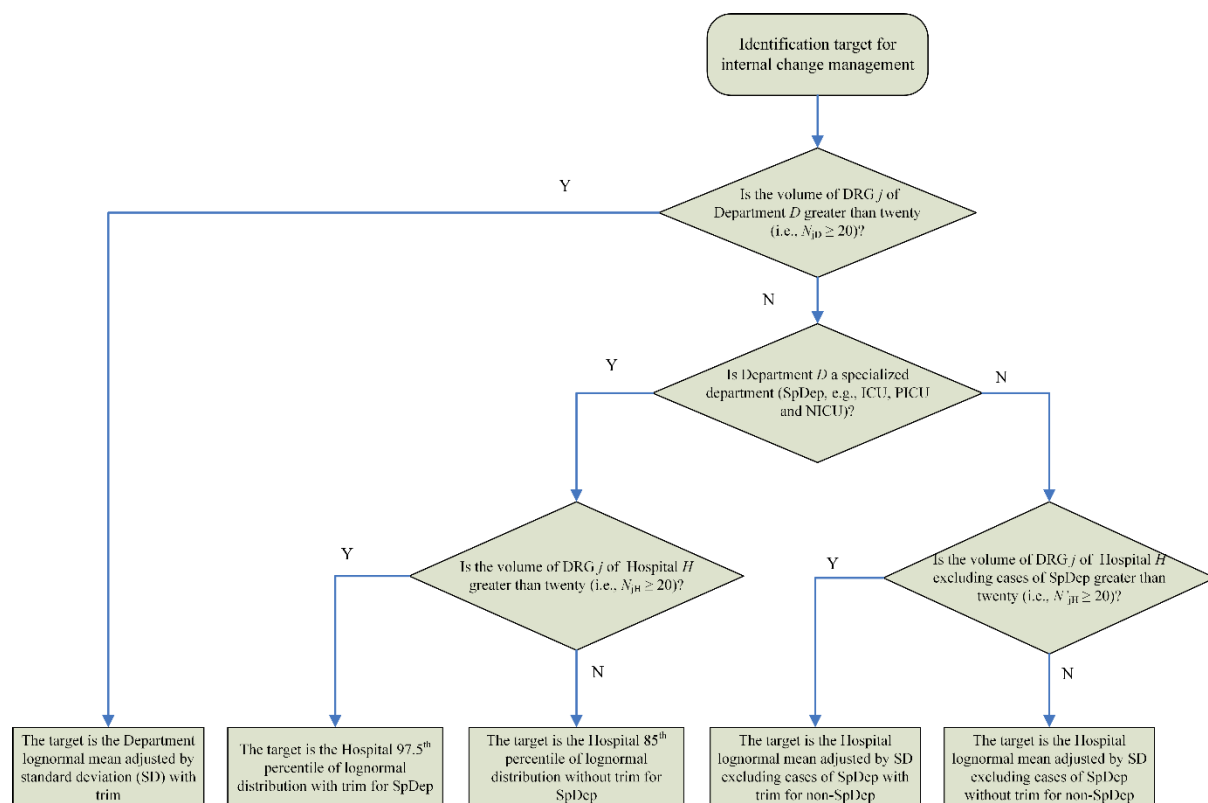
Fig. 2. Target setting at the DRG level for general and specialized departments

For most departments if their DRG volume is greater than 20[2] then the target is set using their lognormal mean adjusted by the standard deviation. Even if the department is a specialised department such as ICU (including PICU, NICU and CCU) provided its volume for DRG j is greater than 20 then the target is set the same as the other departments using the adjusted lognormal mean of the department (3) and standard deviation (4).

The mean of a lognormal distribution is $e^{\widetilde{\mu}_{jD} + \frac{1}{2}\widetilde{\sigma}^2_{jD}}$ and (3) includes the adjustment factor θ.

$$C_j^T = e^{\widetilde{\mu}_{jD} + \frac{1}{2}\widetilde{\sigma}^2_{jD} - \theta\widetilde{\sigma}_{jD}} \qquad (3)$$

---

[2] This selection is based on using 5% for trimming in monthly data but 30 might be more suitable given its reference in the Central Limit Theorem rather in annual data.

$$\sigma_j^T = e^{\tilde{\mu}_{jD} + \frac{1}{2}\tilde{\sigma}_{jD}^2}\sqrt{(e^{\tilde{\sigma}_{jD}^2} - 1)} \tag{4}$$

Where C_j^T, the target of cost for DRG j; $\tilde{\mu}_{jD}$, the mean of the logarithm of DRG j in Department D; $\tilde{\sigma}_{jD}$, the standard deviation of the logarithm of DRG j in Department D; σ_j^T, the target of the standard deviation of DRG j; θ, the adjustment factor.

If the DRG volume is less than 20 for a specialized department such as ICU then the test is whether the hospital volume for the DRG is greater or less than 20. Depending on this the target is set using an appropriate percentile of the Hospital lognormal distribution e.g. 97.5th or 85th percentile.

$$Q_{j(\varphi)} = e^{\tilde{\mu}_{jH} + \tilde{\sigma}_{jH}\Phi(\varphi)} \tag{5}$$

Where $Q_{j(\varphi)}$, the φ-th percentile for DRG j in lognormal distribution; $\Phi(\varphi)$, the φ-th percentile of the standard normal distribution; $\tilde{\mu}_{jH}$, the mean of the logarithm of DRG j in Hospital H; $\tilde{\sigma}_{jH}$, the standard deviation of the logarithm of DRG j in Hospital H.

For non-specialised departments, the test is whether the hospital volume for the DRG is greater or less than 20 excluding specialised department volumes for that DRG. If less than 20 then the target is the hospital lognormal mean adjusted by SD without trim; otherwise the hospital lognormal mean adjusted by SD with trim.

3.2 Evaluation metrics

Cost in this study is an input variable and for the same output, less input is better assuming the same quality. Thus the following metrics are based on cost variances noting that more detailed investigation would involve close scrutiny of inputs making up those variances.

(1) Target Cost Variance

Target Cost Variance (*TCV*) for every DRG equals the variance between its cost target and actual cost. Formulae for the total variances and mean variances are listed in Eq. (6) – (11). If the actual cost is less or equal to the cost target, the *TCV* is positive or 0; otherwise negative.

$$\text{Hospitals:}\qquad TCV_{\text{H}} = \sum_{j_H=1}^{m_H}\sum_{i=1}^{n_{jH}}\left(C_{j_H}^T - C_{ij_H}\right) \tag{6}$$

$$\overline{TCV}_{\text{H}} = \frac{1}{\displaystyle\sum_{j_H=1}^{m_H} n_{j_H}} \sum_{j_H=1}^{m_H}\sum_{i=1}^{n_{jH}}\left(C_{j_H}^T - C_{ij_H}\right) \tag{7}$$

$$\text{Departments:}\qquad TCV_{\text{D}} = \sum_{j_D=1}^{m_D}\sum_{i=1}^{n_{jD}}\left(C_{j_D}^T - C_{ij_D}\right) \tag{8}$$

$$\overline{TCV}_{\text{D}} = \frac{1}{\displaystyle\sum_{j_D=1}^{m_D} n_{j_D}} \sum_{j_D=1}^{m_D}\sum_{i=1}^{n_{jD}}\left(C_{j_D}^T - C_{ij_D}\right) \tag{9}$$

DRG:
$$TCV_j = \sum_{i=1}^{n_j}\left(C_j^T - C_{ij}\right) \tag{10}$$

$$\overline{TCV}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(C_j^T - C_{ij}\right) \tag{11}$$

where $\overline{TCV}_H$, $\overline{TCV}_D$ and $\overline{TCV}_j$ denote the average *TCV* by case for hospitals, departments and DRGs respectively. For hospitals, m_H is the number of DRG groups; n_{j_H}, the number of cases of DRG j_H; C_{ij_H}, the cost of case i of DRG j_H; $C_{j_H}^T$, the cost target of DRG j_H. For departments , m_D is the number of DRG groups; n_{j_D}, number of cases of DRG j_D; C_{ij_D}, the cost of case i of DRG j_D; $C_{j_D}^T$, the cost target of DRG j_D. For DRGs, n_j, the number of cases of DRG j; C_{ij}, the cost of case i of DRG j; C_j^T, the cost target of DRG j.

(2) Target Cost Variance Percent

Since the cost targets for each DRG are different, the mean Target Cost Variance Percent (*TCVP*) evaluates the rate of cost reduction as Eq. (12) – (14). *TCVP* is equal to the ratio of cost reduction to cost target with similar notation as above.

Hospital:
$$\overline{TCVP}_H = \frac{1}{\sum\limits_{j_H=1}^{m_H} n_{j_H}}\sum_{j_H=1}^{m_H}\sum_{i=1}^{n_{j_H}}\left(\frac{C_{j_H}^T - C_{ij_H}}{C_{j_H}^T}\right) \tag{12}$$

Department:
$$\overline{TCVP}_D = \frac{1}{\sum\limits_{j_D=1}^{m_D} n_{j_D}}\sum_{j_D=1}^{m_D}\sum_{i=1}^{n_{j_D}}\left(\frac{C_{j_D}^T - C_{ij_D}}{C_{j_D}^T}\right) \tag{13}$$

DRG:
$$\overline{TCVP}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\left(\frac{C_j^T - C_{ij}}{C_j^T}\right) \tag{14}$$

(3) Normalised Target Cost Variance

The Normalised Target Cost Variance (*NTCV*) is based on the number of standard deviations an observed value is above or below the mean. Z-scores are commonly used in the healthcare sector, e.g., [36], [37], [38].

NTCV is calculated by subtracting the actual cost from the cost target and then dividing the difference by the standard deviation as Eq. (15) – (17).

Hospital:
$$\overline{NTCV}_H = \frac{1}{\sum\limits_{j_H=1}^{m_H} n_{j_H}}\sum_{j_H=1}^{m_H}\sum_{i=1}^{n_{j_H}}\left(\frac{C_{j_H}^T - C_{ij_H}}{\sigma_{j_H}^T}\right) \tag{15}$$

Department:
$$\overline{NTCV}_D = \frac{1}{\sum\limits_{j_D=1}^{m_D} n_{j_D}}\sum_{j_D=1}^{m_D}\sum_{i=1}^{n_{j_D}}\left(\frac{C_{j_D}^T - C_{ij_D}}{\sigma_{j_D}^T}\right) \tag{16}$$

$$\text{DRG:} \qquad \overline{NTCV}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(\frac{C_j^T - C_{ij}}{\sigma_j^T} \right) \qquad\qquad (17)$$

Where $\overline{NTCV}_H$, $\overline{NTCV}_D$ and $\overline{NTCV}_j$, the average Normalised Target Cost Variance (*NTCV*) by cases for hospitals, departments and DRGs respectively; $\sigma_{j_H}^T$, the target of standard deviation target of DRG j_H for hospitals; $\sigma_{j_D}^T$, the target of standard deviation of DRG j_D for departments; σ_j^T, the target of standard deviation of DRG j for DRGs.

4. Empirical application

4.1 Data

Two comprehensive hospitals in the same region are used to illustrate the multi-level system. Hospital A has 1,000 beds and 34 clinical departments, and Hospital B has 950 beds and 32 clinical departments. Hospital A was a new hospital, positioned as an emergency centre, and equipped with advanced medical apparatus and instruments to undertake more emergency functions and responsibilities. Hospital B was an old hospital, positioned as a general hospital to treat common diseases. Both hospitals had implemented Chinese diagnosis-related groups (CN-DRG) developed by the Chinese government with a total of 787 DRGs. All financial data are given in Chinese currency (USD 1=6.5 CNY).

The cost accounting methods employed in both hospitals are based on internationally accepted principles of 'activity-based' cost accounting and cost modelling. Patient clinical and financial data were collected from hospital documentation and finance departments respectively, to identify costs associated with every cost centre/department, and allocated to patients.

There were 2983 and 2585 inpatients reported in Hospital A and B respectively during August 2017. To illustrate the model for the emergency centred hospital and the conventional hospital, six DRGs were randomly selected from the emergency department in each hospital together with the three DRGs with the most cases (OC13, OB23, ES10). This resulted in nine DRGs and 1071 cases from the two hospitals with varied volumes ranging from 6 to 166 to enable the framework to be tested against varying circumstances. Similar departments are assigned the same number for hospitals A and B e.g. A12 stands Department 12 of Hospital A.

The number of cases for Hospital A (605) and Hospital B (466) for their respective departments, and the explanations of these DRGs are presented in Table 1. Nine departments are the same across both hospitals except for A12 (PICU) and B30 (Infections). Note that these volumes for one month only and display considerable variability in volumes. The regional, hospital and departmental cost means and standard deviations for these DRGs are reported in Table 2.

Table 1 Number of cases distributed by DRG and department and DRG description

	Dep. Name	BB23	BK19	BR23	BY13	ES10	ES13	FB29	OB23	OC13	Cases	DRGs
	Hospital A	14	15	74	16	107	20	80	113	166	605	9
A2	ICU	2	5		3		1				11	4
A4	Surgical Oncology						1				1	1
A6	Obstetrics								113	166	279	2
A12	PICU					24					24	1
A13	Neurosurgery	12	8	3	13						36	4
A14	Pediatric					83					83	1
A19	Respiratory		2				14				16	2
A20	Neurology			61							61	1
A21	Cardiovasology			3			2	79			84	3
A26	Geriatrics			7			2	1			10	3
	Hospital B	12	6	122	19	99	65	34	62	47	466	9
B2	ICU	2		9	2		2				15	4
B4	Surgical Oncology		2								2	1
B6	Obstetrics								62	47	109	2
B13	Neurosurgery	10	2		17						29	3
B14	Pediatric					99					99	1
B19	Respiratory			1			51				52	2
B20	Neurology			89							89	1
B21	Cardiovasology			6				24			30	2
B26	Geriatrics		1	17			5	10			33	4
B30	Infections		1				7				8	2

DRG	Description	Remark
BB23	Other craniotomies, w CC	w=with
BK19	Neurological diagnosis with ventilator support	CC=complication and comorbidity conditions.
BR23	Cerebral ischemic disease, w CC	
BY13	Intracranial injury, w CC	
ES10	Respiratory infection/inflammation, age<17	
ES13	Respiratory infection/inflammation, w CC	
FB29	Cardiac valvular surgery with cardiac catheterization	
OB23	Cesarean section, w CC	
OC13	Vaginal delivery with surgery, w CC	

Table 2 Descriptive statistics

| DRG | Region Level | | Hospital Level | | | | | | Departmental Level | | | | | |
---	Mean	SD	Cases of Hosp. A	Mean of Hosp. A	SD of Hosp. A	Cases of Hosp. B	Mean Of Hosp. B	SD of Hosp. B	Dep.of Hosp. A	Cases	Mean	Dep. of Hosp. B	Cases	Mean
BB23	95935	46,690	14	90,540	62,073	12	108,049	58,728	A2	11	83,549	B2	15	41,164
BK19	50793	44,674	15	52,741	40,772	6	43,739	17,866	A4	1	7,267	B4	2	60,835
BR23	9341	6,850	74	11,578	5,441	122	8,978	4,573	A6	279	5,497	B6	109	6,728
BY13	11887	16,748	16	7,626	4,169	19	10,208	7,922	A12	24	9,483	B13	29	38,632
ES10	4533	3,768	107	5,044	3,080	99	3,669	1,803	A13	36	35,917	B14	99	3,669
ES13	9037	8,161	20	9,909	7,747	65	7,202	3,034	A14	83	3,761	B19	52	6,691
FB29	15273	14,309	80	13,587	14,865	34	16,258	14,474	A19	16	7,378	B20	89	8,435
OB23	8894	2,836	113	8,749	1,895	62	8,896	1,719	A20	61	12,559	B21	30	15,835
OC13	3706	1,331	166	3,283	993	47	3,867	1,192	A21	84	13,406	B26	33	10,147
SUM	–	–	605	–	–	466	–	–	A26	10	9,309	B30	8	12,661

4.2 Top down evaluation using regional benchmarks

To test the differences between the region and the two hospitals, the tests for each DRG are reported in Table 3.

Table 3 Variance and Mean Test for each DRG cost between Hospital A and B

DRG	Hosp. A		Hosp. B		Hosp. A vs Hosp. B	
	χ^2-test (p value)	Z-test (p value)	χ2-test (p value)	Z-test (p value)	F-test (p value)	t-test (p value)
BB23	0.042	0.049	0.097	0.002	0.27	0.47
BK19	0.367	0.866	0.023	0.699	0.093	0.515
BR23	0.006	0.005	0	0.558	0.015	0.001
BY13	0	0.309	0	0.662	0.034	0.228
ES10	0.003	0.16	0	0.023	0	0
ES13	0.418	0.633	0	0.07	0	0.142
FB29	0.377	0.117	0.358	0.995	0.115	0.378
OB23	0	0.587	0	0.995	0.893	0.612
OC13	0	0	0.171	0.405	0.405	0.001

- *Z*-test for comparing Regional mean cost with Hospital A mean cost and Hospital B mean cost;

- χ^2-test for comparing Regional standard deviation with Hospital A standard deviation and Hospital B standard deviation;

- *T*-test for comparing Hospital A mean cost with Hospital B mean cost;

- *F*-test for comparing Hospital A standard deviation with Hospital B standard deviation.

The *Z*-test and χ^2–test identify any significant differences between each hospital and the region highlighting shortfalls or surpluses in funding.

At the hospital level, the regional mean (1) is set as the target for the evaluation and Table 4 reports favourable benchmark variances 95,962 and 136,694 for both hospitals suggesting satisfactory performance relative to external reference prices. However, the summary metric scores of TCV_H, $\overline{TCV_H}$, $\overline{TCVP_H}$, and $\overline{NTCV_H}$ suggest that Hospital B performed at a higher level than Hospital A, even though the average cost of Hospital B (10,656) was slightly higher than that of Hospital A (10,572). As an emergency hospital, Hospital A has a different structure to Hospital B that aligns with other studies [39,40].

Since the incomes of hospitals are primarily from the DRG reimbursement under DRG-PPS, Column 3 and 7 of Table 4 show which DRGs are profitable and which DRGs are in deficit. For example, BB23, BY13, FB29, OB23 and OC13 were profitable while BK19, BR23, ES10 and ES13 were in deficit in Hospital A; BK19, BR23, BY13, ES10 and ES13 were profitable while BB23, FB29, OB23 and OC13 were in deficit in Hospital B.

The individual DRG metrics show considerable differences. BB23 has positive metrics for hospital A but not for hospital B. The *TCV* for A (B) is 75,523 (-145,374) overall equating to

5,395 (-12,114) per case, 5.6% (-12.6%) of target cost and 0.116 (-0.259) standardised score in terms of variability. Each metric tells a slightly different story depending on whether the focus is on case profitability, the proportion of the reimbursement price or variability from the reimbursement price. For example, management of hospital A might be concerned that BR23 is losing them over 2,000 yen per case which is 24% of the price and almost one third of a standard deviation above target.

Table 4 reports the performance metrics using regional benchmark, in both hospitals ICU Department 2 has the highest unfavourable results and Department 13 the most favourable. Some insight is provided for this in the last 10 columns (department level) where Departments 2 and 13 have opposite performance metrics reflecting differences in the complexity of the patients they serve.

Using the regional mean and standard deviation as the benchmarks can reflect the performance of hospitals from a regional perspective, but becomes more difficult to evaluate departmental internal performance. As noted large deficits are reported for department 2 (ICU) with opposite profitable measures for department 13 (Neurosurgery). It is noteworthy that the mean for ICU lies more than two standard deviations from the regional mean in both hospitals suggesting that the regional target is unrealistic for this department. While this might provide some insight for hospital management, it is likely that departmental managers will either be irritated (ICU) or complacent (Neurosurgery). From a change management perspective, different targets are needed that consider the characteristics of the hospital and its internal structures.

Table 4 Performance evaluation using regional benchmarks

DRG	Region Benchmark	Hosp. A TCV_H	$\overline{TCV}_H$	$\overline{TCVP}_H$	$\overline{NTCV}_H$	Hosp. B TCV_H	$\overline{TCV}_H$	$\overline{TCVP}_H$	$\overline{NTCV}_H$	Dep.	Departments of Hosp. A TCV_D	$\overline{TCV}_D$	$\overline{TCVP}_D$	$\overline{NTCV}_D$	Dep.	Departments of Hosp. B TCV_D	$\overline{TCV}_D$	$\overline{TCVP}_D$	$\overline{NTCV}_D$
BB23	95935	75,523	5,395	.056	.116	(145,374)	(12,114)	(.126)	(.259)	A2	(428,509)	(38,955)	(.810)	(1.074)	B2	(299,673)	(19,978)	(.846)	(1.126)
BK19	50793	(29,221)	(1,948)	(.038)	(.044)	42,324	7,054	.139	.158	A4	1,770	1,770	.196	.217	B4	(20,084)	(10,042)	(.198)	(.225)
BR23	9341	(165,551)	(2,237)	(.240)	(.327)	44,266	363	.039	.053	A6	86,625	310	.075	.210	B6	(7,705)	(71)	(.019)	(.053)
BY13	11887	68,175	4,261	.358	.254	31,892	1,679	.141	.100	A12	(118,806)	(4,950)	(1.092)	(1.314)	B13	142,698	4,921	.216	.186
ES10	4533	(54,723)	(511)	(.113)	(.136)	85,526	864	.191	.229	A13	447,111	12,420	.312	.359	B14	85,526	864	.191	.229
ES13	9037	(19,723)	(986)	(.109)	(.121)	119,256	1,835	.203	.225	A14	64,083	772	.170	.205	B19	122,306	2,352	.260	.290
FB29	15273	134,856	1,686	.110	.118	(33,492)	(985)	(.064)	(.069)	A19	107,768	6,736	.308	.344	B20	80,651	906	.097	.132
OB23	8894	16,407	145	.016	.051	(136)	(2)	.000	(.001)	A20	(196,286)	(3,218)	(.344)	(.470)	B21	(52,438)	(1,748)	(.091)	(.080)
OC13	3706	70,219	423	.114	.318	(7,569)	(161)	(.043)	(.121)	A21	126,561	1,507	.098	.109	B26	72,646	2,201	.131	.151
SUM	-	95,962	159	.006	.053	136,694	293	.074	.076	A26	5,645	564	.018	.056	B30	12,768	1,596	.108	.120

4.3 Bottom up evaluation using change management targets

Following the bottom up approach depicted in Figure 1, targets are set for DRGs contextualised by the department providing them. Depending on volumes, either the adjusted lognormal mean (3) and standard deviation (4) or percentiles (5) are used to set the targets as set out in Figure 2. The adjustment factor θ will be further discussed in Section 5, but for explanatory purposes, a moderate θ of 0.03 is used for this Section. Table 5 displays the targets for each DRG and each department.

Table 5 Target settings for each DRG and each department in change management

Dep.	BB23	BK19	BR23	BY13	ES10	ES13	FB29	OB23	OC13
A2	173,235	91,648		11,675		23,525			
A4						8,565			
A6								8,502	3,134
A12					9,309				
A13	79,565	32,375	11,270	6,496					
A14					3,624				
A19		32,375				8,565			
A20			10,206						
A21			11,270			8,565	11,702		
A26			11,270			8,565	11,497		
B2	172,114		16,058	18,319		9,750			
B4		43,822							
B6								8,664	3,706
B13	92,789	43,822		8,604					
B14					3,472				
B19			8,019			6,575			
B20			8,089						
B21			8,019				15,488		
B26		43,822	8,019			6,634	14,169		
B30		43,822				6,575			

Table 6 shows the results of departments in Hospital A and B using change management target. Note that the targets for departments 2 (ICU) and 12 (PICU in hospital A) are closer to actual cost with commensurate lower unfavourable metrics. The targets of ICUs with volume less than twenty are calculated using the percentiles from (5) while that of ICUs with volume greater than twenty using adjusted lognormal mean from (3) in terms of the logic of Figure 2.

In contrast, the benchmarking analysis reported in Table 4 showed the ICUs (Dep.2 and Dep. 12) having the biggest loss and worst performance in both hospitals A and B. This is likely to be unfair since ICU is a special department that receives patients with severe or life-threatening illnesses and injuries, requiring constant care, close supervision of life support equipment and medication to support the delivery of both acute care and complex elective surgery. ICUs are staffed by highly trained physicians, nurses and respiratory therapists who specialize in caring for critically ill patients. They are also distinguished from general hospital wards by a higher staff-to-patient ratio and access to advanced medical resources and equipment that are not routinely available elsewhere. Typically, the costs of ICUs are much higher than general hospital wards.

In our view, Table 5 provides a more balanced target that takes into consideration the special

characteristics of these specialised departments. Figure 2 was used to determine the targets for the departmental DRGs so DRG BB23 and BY13 for ICU in both hospitals were calculated using the 85th percentile. In contrast for ES13, the 97.5th percentile of trimmed hospital cases is used to set the target for Department 2 in both hospitals.

Table 6 Performance evaluation using internal change target

	Departments of Hosp. A					Departments of Hosp. B				DRG	Hosp. A				Hosp. B			
Dep.	TCV_D	$\overline{TCV}_D$	$\overline{TCVP}_D$	$\overline{NTCV}_D$	**Dep.**	TCV_D	$\overline{TCV}_D$	$\overline{TCVP}_D$	$\overline{NTCV}_D$		TCV_H	$\overline{TCV}_H$	$\overline{TCVP}_H$	$\overline{NTCV}_H$	TCV_H	$\overline{TCV}_H$	$\overline{TCVP}_H$	$\overline{NTCV}_H$
A2	(55,784)	(5,071)	(.077)	(.227)	**B2**	(72,575)	(4,838)	(.152)	(.499)	**BB23**	33,680	2,406	.047	.033	(24,474)	(2,039)	(.005)	(.022)
A4	1,298	1,298	.152	.205	**B4**	(34,025)	(17,013)	(.388)	(.906)	**BK19**	(9,130)	(609)	(.004)	(.009)	501	83	.002	.004
A6	(52,628)	(189)	(.040)	(.268)	**B6**	(21,976)	(202)	(.044)	(.250)	**BR23**	(87,715)	(1,185)	(.123)	(.385)	(38,475)	(315)	(.035)	(.104)
A12	(4,180)	(174)	(.019)	(.065)	**B13**	41,485	1,431	.040	.067	**BY13**	(2,545)	(159)	(.017)	(.035)	(11,057)	(582)	(.017)	(.042)
A13	39,025	1,084	.013	.029	**B14**	(19,531)	(197)	(.057)	(.157)	**ES10**	(15,505)	(145)	(.033)	(.128)	(19,531)	(197)	(.057)	(.157)
A14	(11,325)	(136)	(.038)	(.146)	**B19**	(4,562)	(88)	(.015)	(.044)	**ES13**	(11,925)	(596)	(.036)	(.103)	(34,113)	(525)	(.072)	(.223)
A19	66,603	4,163	.271	.363	**B20**	(30,811)	(346)	(.043)	(.116)	**FB29**	(150,997)	(1,887)	(.161)	(.611)	(39,359)	(1,158)	(.070)	(.096)
A20	(143,532)	(2,353)	(.231)	(.644)	**B21**	(55,199)	(1,840)	(.101)	(.126)	**OB23**	(27,901)	(247)	(.029)	(.242)	(14,388)	(232)	(.027)	(.237)
A21	(150,673)	(1,794)	(.157)	(.586)	**B26**	20,145	610	(.019)	(.097)	**OC13**	(24,727)	(149)	(.048)	(.286)	(7,588)	(161)	(.067)	(.266)
A26	14,430	1,443	.076	.430	**B30**	(11,435)	(1,429)	(.202)	(.571)	**SUM**	(296,765)	(491)	(.061)	(.278)	(188,485)	(404)	(.048)	(.159)

Comparisons between the targets for benchmarking and change management are reported in the appendix. The change management targets are generally more demanding than the regional targets used for benchmarking and both hospitals now show total unfavourable variances in contrast to the favourable variances. While some individual DRGs have improved variances such as BK19 and BR23 for hospital A, others such as FB29, OB23 and OC13 show worse performance compared with regional benchmarking. At the hospital level, change management targets are lower than regional prices which may or may not be acceptable depending on management strategy and incentive schemes.

5. Discussion and implications
5.1 Target setting

Cost target setting is critical to cost containment because it allows all medical staff to know cost upper limits with appropriate accountability to investigate deviations from target. A cost target that is easy to achieve may result in inefficiencies. If it is too difficult to achieve, this can have a negative impact on motivation and performance.

For benchmarking, the regional DRG mean and standard deviation are used as the target benchmark, which helps management identify areas of strength and weakness and inform the process of target setting for change management or target costing.

At Department and DRG levels, targets need to be attainable as well as being reasonably tight, i.e. a department manager can achieve target a certain percentage of the time. For general departments we use the adjusted lognormal mean with trimming, but for ICUs with low volumes we use the lognormal percentile.

Our example was for only one month so on an annual basis which would be the normal target setting period, volumes are likely to be much higher. Thus, even if a department had only 5 cases for a DRG monthly, this would be 60 for a whole year. Notwithstanding, there may be some DRGs where the annual volume for a specialised department is less than, say, 20 but the annual volume at the hospital level is well above that. The question is where the mean of the distribution for that department lies in terms of the hospital distribution? That is an empirical question which we could answer if we had more monthly data but for this illustration we have assumed 97.5th percentile rather arbitrarily based on a 5% outlier decision.

The relationship between regional prices and internal targets provides a good example of linking an open with a closed system. Table 7 reports the different variances depending on which approach and target is selected. Benchmarking targets using regional prices have different variances from those using change management internally based targets. However, the more interesting variance is between the regional benchmarking target and the change management target of 395,010 and 325,178 for hospitals A and B respectively. These are 6.1% and 6.4% of the regional benchmarking target.

Table 7. Hospital level total variances

	Hospital Total Targets	
	Hosp. A	Hosp. B
Regional Benchmarking (1)	6,494,240	5,102,497
Change Management Target (2)	6,099,230	4,777,319
(1)-(2)	395,010	325,178
Actual Cost (3)	6,395,996	4,965,803
Regional Benchmarking Variance (1)-(3)	98,244	136,694
Change Management Variance (2)-(3)	(296,765)	(188,485)

These variances and percentages may be considered too high and although there are several ways to address this, an obvious avenue would be to examine the value of the adjustment factor θ and percentile φ.

Table 8 sets out 4 scenarios depending on the position of the hospital mean and standard deviation relative to the region. Some examples of possible values for θ are provided illustrating a lenient or aggressive approach.

Table 8. Suggested adjustment factor θ considering the relationship between regional prices and internal targets

Scenario	Hosp. Mean	Hosp. SD	θ	
			Lenient	Aggressive
1	< Regional Mean	< Regional SD	0~0.01	0.01~0.02
2	< Regional Mean	> Regional SD	0.01~0.03	0.03~0.05
3	> Regional Mean	< Regional SD	0.02~0.03	0.04~0.07
4	> Regional Mean	> Regional SD	0.03~0.05	0.05~0.1

An appropriate percentile φ is an empirical question that requires examination of a low volume DRG for ICU relative to the hospital wide costs. We argue that ICU cases are more complex and therefore these costs will lie around the higher end of the distribution at the hospital level.

In summary, target setting for internal performance evaluation should consider the characteristics of the hospital, departments and DRGs, to prevent under-treatment or disincentives due to one size fits all, and enable cost containment.

The *TCV* and $\overline{TCV}$ are intuitive approaches to gauge cost containment performance using the variances between the target and the actual cost but they may sometimes obscure performance. For example, if DRG α has a cost target $5000, and actual costs for a department are $4800, then the department is rewarded with 200 points. DRG β's cost target may be $500, and its actual costs for another department $480, resulting in 20 points for that department. Reducing DRG α costs by $20 may be relatively easy compared with DRG β, which motivates the $\overline{TCVP}$ metric. In this example, the $\overline{TCVP}$ s of both DRGs are 0.04. This implies the same level of effort in cost containment by both departments. Variability is also important since a

high (low) variance may imply greater (less) ability to better manage costs. The $\overline{NCTV}$ metric addresses this issue and for benchmarking using regional targets, can clearly indicate how appropriate the target is for specific departments such as ICU.

5.2 Performance management

Two sets of targets might be considered overly complicated. However, using regional targets for both benchmarking and internal change management can be problematic. Schreyögg et al [22] describe cost compression in Germany where the prices of high cost DRGs were set low relative to their actual costs whereas the prices of low cost DRGs were set high relative to their costs. Theoretically, DRG-based payments provide strong incentives to increase the number of cases treated and to reduce the number of services per case. In contrast to fee-for-service systems, DRGs encourage hospitals to limit their activities to necessary services; in contrast to external budgets, DRGs encourage hospitals to treat more patients. In terms of expenditure control, the effect of DRG-based payments depends on which effect dominates: increasing the number of cases or decreasing the number of services per case. If the DRG does not adequately control for differences between patient populations or differences in the services provided within the DRG, payments for highly complex cases will be too low and payments for less complex cases will be too high [1].

Some researchers argue that a sole emphasis on costs may reduce the quality of medical care by shifting emphasis away from providing quality [41,42]. Many countries are experimenting with new methods to realign payment incentives in health care to encourage higher quality and more efficient care by pay-for-performance (P4P) [43,44]. These P4P schemes are testing whether new ways of paying providers (i.e. hospitals, primary care, integrated systems), which include a synthetic measure of quality, show improvements in the quality of care and value for money in health [45]. The model proposed in this study can be applied in P4P to stimulate external and internal competition among healthcare providers and to motivate the search for lower costs while maintaining volumes and quality. It is particularly suited to healthcare organisations facing external funding processes while also having to provide quality services to varying levels of complexity for some DRGs.

6. Conclusion, limitations, and future research

There is a danger that hospitals engage in 'cream-skimming'[46]; that is, they attempt to admit only those patients within each DRG that can be expected to have costs below the payment rate (for example, by selecting patients without co-morbidities, if these are not adequately accounted for in the DRG system) or that they 'dump' unprofitable patients by transferring them to other providers or avoiding them altogether [47]. The promise of DRG systems had to be weighed against their weaknesses and risks, including that DRG systems create an incentive for hospitals to skimp on services provided per admission (that is, to undertreat), discharge patients prematurely, and cherry-pick low-cost patients. The costs of developing new systems to prevent gaming of DRG payments (for example, through upcoding and splitting of admissions) also needed to be considered [48]. If DRG systems adequately account for differences between patients (by considering all relevant procedures), the incentives for certain unintended consequences, such as cream-skimming and skimping/undertreatment, could be greatly reduced.

The multi-level system assesses cost containment performance from various perspectives, enabling high-level variances to be decomposed into departmental and DRG variances thereby identifying sources of performance variability and financial impact. The perspectives range from macro to micro, external to internal, providing accountability throughout the organisation. As such it can provide a valuable system under PPS and any regime where external DRG prices are available. To avoid undertreatment, high cost targets are allowed in terms of their

probability distributions for special departments (e.g., ICU); to avoid cream-skimming, set reasonable cost targets for non-special sectors in light of their probability distributions. Note that for both special departments and non-special departments, θ and φ are adjusted relying on their probability distributions to ensure that they operate within a reasonable range.

There are several limitations to this research. First, the research was conducted using two Chinese hospitals belonging to one region. Future research in different regions / countries can explore the generalizability of the multi-level model. Second, excluding outliers was argued to provide a better approximation of the regional mean in setting targets for DRGs within departments. A testable hypothesis would be that the mean cost or LOS of a DRG (e.g. hip replacement) should be the same across similar departments (e.g. orthopaedic) in different hospitals. Third, low volumes for some DRGs in specialised departments are challenging when setting appropriate targets. This study used percentiles at the high side of the hospital distribution of costs on the argument that complex cases populate this area. Research across similar departments in different hospitals could confirm this and identify what the appropriate percentiles should be. Forth, value assessments have become a critical foundation for policymaking, helping to create better-informed pricing. Some rely on the conventional cost-effectiveness analysis (CEA) metric of the quality-adjusted life-year (QALY), whereas better-informed pricing can lead to higher-value spending, which in turn can improve health outcomes, strengthen health system sustainability, and promote high-value innovation.

Finally, the model has a strong cybernetic flavour with the combining of both open and closed systems with intervening variables θ and percentile φ. More formal cybernetic modelling could provide new directions in performance evaluation systems in both healthcare and other organisations.

Appendix
Table A.1 Comparison at the Hospital Level breakdown into DRGs

Hosp.	DRG	Cases	Mean	Regional prices as benchmark					Internal Prices for Change Management				
				Target	TCV_j	$\overline{TCV}_j$	$\overline{TCVP}_j$	$\overline{NTCV}_j$	Target	TCV_j	$\overline{TCV}_j$	$\overline{TCVP}_j$	$\overline{NTCV}_j$
A	BB23	14	90,540	95,935	75,523	5,395	.056	.116	92,946	33,680	2,406	.047	.033
	BK19	15	52,741	50,793	(29,221)	(1,948)	(.038)	(.044)	52,132	(9,130)	(609)	(.004)	(.009)
	BR23	74	11,578	9,341	(165,551)	(2,237)	(.240)	(.327)	10,393	(87,715)	(1,185)	(.123)	(.385)
	BY13	16	7,626	11,887	68,175	4,261	.358	.254	7,467	(2,545)	(159)	(.017)	(.035)
	ES10	107	5,044	4,533	(54,723)	(511)	(.113)	(.136)	4,900	(15,505)	(145)	(.033)	(.128)
	ES13	20	9,909	9,037	(19,723)	(986)	(.109)	(.121)	9,313	(11,925)	(596)	(.036)	(.103)
	FB29	80	13,587	15,273	134,856	1,686	.110	.118	11,700	(150,997)	(1,887)	(.161)	(.611)
	OB23	113	8,749	8,894	16,407	145	.016	.051	8,502	(27,901)	(247)	(.029)	(.242)
	OC13	166	3,283	3,706	70,219	423	.114	.318	3,134	(24,727)	(149)	(.048)	(.286)
	SUM	605	10,572	10,734	95,962	159	.006	.053	10081	(296,765)	(491)	(.061)	(.278)
B	BB23	12	108,049	95,935	(145,374)	(12,114)	(.126)	(.259)	106,010	(24,474)	(2,039)	(.005)	(.022)
	BK19	6	43,739	50,793	42,324	7,054	.139	.158	43,822	501	83	.002	.004
	BR23	122	8,978	9,341	44,266	363	.039	.053	8,663	(38,475)	(315)	(.035)	(.104)
	BY13	19	10,208	11,887	31,892	1,679	.141	.100	9,626	(11,057)	(582)	(.017)	(.042)
	ES10	99	3,669	4,533	85,526	864	.191	.229	3,472	(19,531)	(197)	(.057)	(.157)
	ES13	65	7,202	9,037	119,256	1,835	.203	.225	6,677	(34,113)	(525)	(.072)	(.223)
	FB29	34	16,258	15,273	(33,492)	(985)	(.064)	(.069)	15,100	(39,359)	(1,158)	(.070)	(.096)
	OB23	62	8,896	8,894	(136)	(2)	.000	(.001)	8,664	(14,388)	(232)	(.027)	(.237)
	OC13	47	3,867	3,706	(7,569)	(161)	(.043)	(.121)	3,706	(7,588)	(161)	(.067)	(.266)
	SUM	466	10,656	10,950	136,694	293	.074	.076	10252	(188,485)	(404)	(.048)	(.159)

References

1. Busse, R.; Geissler, A.; Aaviksoo, A.; Cots, F.; Häkkinen, U.; Kobel, C.; Mateus, C.; Or, Z.; O'Reilly, J.; Serdén, L. Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals? *BMJ* **2013**, *346*, f3197, https://doi.org/10.1136/bmj.f3197.

2. Fetter, R.B. Diagnosis related groups: understanding hospital performance. *Interfaces* **1991**, *21*, 6-26, https://doi.org/10.1287/inte.21.1.6.

3. Quentin, W.; Geissler, A.; Scheller-Kreinsen, D.; Busse, R. Understanding DRGs and DRG-based hospital payment in Europe. In *Diagnosis-Related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*, Open University Press and Mc Graw Hill: Berkshire, England, 2011; pp. 23-36.

4. Langenbrunner, J.C.; Wiley, M.M. Hospital payment mechanisms: theory and practice in transition countries. In *Hospitals in a changing Europe*, McKee, M., Healy, J., Eds. Open University Press: Buckingham, Philadelphia, 2002; pp. 150-176.

5. Burwell, S.M. Setting value-based payment goals—HHS efforts to improve US health care. *N. Engl. J. Med.* **2015**, *372*, 897-899, https://doi.org/10.1056/NEJMp1500445.

6. Kairies-Schwarz, N.; Souček, C. Performance Pay in Hospitals: An Experiment on Bonus–Malus Incentives. *Int. J. Environ. Res. Public Health* **2020**, *17*, https://doi.org/10.3390/ijerph17228320.

7. Liu, X.; Xu, J.; Yuan, B.; Ma, X.; Fang, H.; Meng, Q. Containing medical expenditure: lessons from reform of Beijing public hospitals. *BMJ* **2019**, *365*, l2369, https://doi.org/10.1136/bmj.l2369.

8. Shon, C.; You, M. Evaluation of Health Policy Governance in the Introduction of the New DRG-Based Hospital Payment System from Interviews with Policy Elites in South Korea. *Int. J. Environ. Res. Public Health* **2020**, *17*, https://doi.org/10.3390/ijerph17113757.

9. HE, A.J. Is the Chinese health bureaucracy incapable of leading healthcare reforms?: the case of Fujian Province. *China: An International Journal* **2012**, *10*, 93-112, https://doi.org/10.1353/chn.2012.0003.

10. Jian, W.; Huang, Y.; Hu, M.; Zhang, X. Performance evaluation of inpatient service in Beijing: a horizontal comparison with risk adjustment based on Diagnosis Related Groups. *BMC Health Serv. Res.* **2009**, *9*, 72-80, https://doi.org/10.1186/1472-6963-9-72.

11. Park, H.; Shin, Y. Measuring Case-Mix Complexity of Tertiary Care Hospitals Using DRGs. *Health Care Manag. Sci.* **2004**, *7*, 51-61, https://doi.org/10.1023/B:HCMS.0000005398.52789.6d.

12. Ji, X.; Fang, Y.; Liu, J. Performance assessment of the inpatient medical services of a clinical subspecialty: A case study with risk adjustment based on diagnosis-related groups in China. *Medicine* **2018**, *97*, 1-7, https://doi.org/10.1097/MD.0000000000010855.

13. Stock, G.N.; McDermott, C.; Anand, G. Average surgeon-level volume and hospital performance. *International Journal of Production Economics* **2016**, *182*, 253-262.

14. Bichescu, B.C.; Bradley, R.V.; Smith, A.L.; Wei, W. Benefits and implications of competing on process excellence: Evidence from California hospitals. *International Journal of Production Economics* **2018**, *202*, 59-68, https://doi.org/10.1016/j.ijpe.2018.05.013.

15. Nathanson, M. DRG cost-per-case management. Comprehensive cost accounting systems give chains an edge. *Mod. Healthc.* **1984**, *14*, 122, 124, 128.

16. Raulinajtys-Grzybek, M. Cost accounting models used for price-setting of health services: an international review. *Health Policy* **2014**, *118*, 341-353, https://doi.org/10.1016/j.healthpol.2014.07.007

17. Tan, S.S.; Geissler, A.; Serdén, L.; Heurgren, M.; Van Ineveld, B.M.; Redekop, W.K.; Hakkaart-van Roijen, L. DRG systems in Europe: variations in cost accounting systems among 12 countries. *The European Journal of Public Health* **2014**, *24*, 1023-1028, https://doi.org/10.1093/eurpub/cku025

18. Ridderstolpe, L.; Johansson, A.; Skau, T.; Rutberg, H.; Åhlfeldt, H. Clinical process analysis and activity-based costing at a heart center. *J. Med. Syst.* **2002**, *26*, 309-322, https://doi.org/10.1023/A:1015816703951.

19. Tan, S.S.; Rutten, F.; Van Ineveld, B.; Redekop, W.; Hakkaart-van Roijen, L. Comparing methodologies for the cost estimation of hospital services. *The European Journal of Health Economics* **2009**, *10*, 39-45, https://doi.org/10.1007/s10198-008-0101-x

20. Kuntz, L.; Scholtes, S.; Vera, A. DRG cost weight volatility and hospital performance. *OR Spectrum* **2008**, *30*, 331-354, https://doi.org/10.1007/s00291-006-0063-2

21.   Ghaffari, S.; Doran, C.; Wilson, A.; Aisbett, C.; Jackson, T. Investigating DRG cost weights for hospitals in middle income countries. *Int. J. Health Plann. Mgmt.* **2009**, *24*, 251-264, https://doi.org/10.1002/hpm.948.

22.   Schreyögg, J.; Tiemann, O.; Busse, R. Cost accounting to determine prices: how well do prices reflect costs in the German DRG-system? *Health Care Manag. Sci.* **2006**, *9*, 269-279, https://doi.org/10.1007/s10729-006-9094-0

23.   Phelan, P.D.; Tate, R.; Webster, F.; Marshall, R.P. DRG cost weights — getting it right. *Medical Journal of Australia* **1998**, *169*, S36-S38, https://doi.org/10.5694/j.1326-5377.1998.tb123474.x.

24.   Suthummanon, S.; Omachounu, V.K. DRG-Based Cost Minimization Models: Applications in a Hospital Environment. *Health Care Manag. Sci.* **2004**, *7*, 197-205, https://doi.org/10.1023/B:HCMS.0000039382.05904.c3

25.   Boes, S.; Napierala, C. Assessment of the introduction of DRG-based reimbursement in Switzerland: Evidence on the short-term effects on length of stay compliance in university hospitals. *Health Policy* **2021**, https://doi.org/10.1016/j.healthpol.2021.01.010.

26.   Polverejan, E.; Gardiner, J.C.; Bradley, C.J.; Holmes‐Rovner, M.; Rovner, D. Estimating mean hospital cost as a function of length of stay and patient characteristics. *Health Econ.* **2003**, *12*, 935-947, https://doi.org/10.1002/hec.774

27.   Skeie, B.; Mishra, V.; Vaaler, S.; Amlie, E. A comparison of actual cost, DRG‐based cost, and hospital reimbursement for liver transplant patients. *Transpl. Int.* **2002**, *15*, 439-445.

28.   Kaier, K.; Wolkewitz, M.; Hehn, P.; Mutters, N.T.; Heister, T. The impact of hospital-acquired infections on the patient-level reimbursement-cost relationship in a DRG-based hospital payment system. *International Journal of Health Economics and Management* **2020**, *20*, 1-11, https://doi.org/10.1007/s10754-019-09267-w.

29.   Emmanuel, C.; Otley, D.; Merchant, K. *Accounting for Management Control*, 2nd edition ed.; Chapman and Hall: London, 1990.

30.   Cots, F.; Elvira, D.; Castells, X.; Sáez, M. Relevance of Outlier Cases in Case Mix Systems and Evaluation of Trimming Methods. *Health Care Manag. Sci.* **2003**, *6*, 27-35, https://doi.org/10.1023/A:1021908220013.

31.   Kraus, T.W.; Mieth, M.; Schneider, T.; Farrenkopf, I.; Mehrabi, A.; Schemmer, P.; Encke, J.; Sauer, P.; Büchler, M.W. Cost distribution of orthotopic liver transplantation: single-center analysis under DRG-based reimbursement. *Transplantation* **2005**, *80*, S97-S100, https://doi.org/10.1097/01.tp.0000186939.35230.fe

32.   Pirson, M.; Dramaix, M.; Leclercq, P.; Jackson, T. Analysis of cost outliers within APR-DRGs in a Belgian general hospital: two complementary approaches. *Health policy* **2006**, *76*, 13-25, https://doi.org/10.1016/j.healthpol.2005.04.008

33.   Briggs, A.; Gray, A. The distribution of health care costs and their statistical analysis for economic evaluation. *J. Health Serv. Res. Policy* **1998**, *3*, 233-245, https://doi.org/10.1177/135581969800300410

34.   Cots, F.; Mercadé, L.; Castells, X.; Salvador, X. Relationship between hospital structural level and length of stay outliers: Implications for hospital payment systems. *Health policy* **2004**, *68*, 159-168, https://doi.org/10.1016/j.healthpol.2003.09.004

35.   Palmer, G.; Reid, B. Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. *Health services management research* **2001**, *14*, 71-81, https://doi.org/10.1258/0951484011912564

36.   Austin, J.M.; D'Andrea, G.; Birkmeyer, J.D.; Leape, L.L.; Milstein, A.; Pronovost, P.J.; Romano, P.S.; Singer, S.J.; Vogus, T.J.; Wachter, R.M. Safety in numbers: the development of Leapfrog's composite patient safety score for US hospitals. *Journal of patient safety* **2014**, *10*, 64-71, https://doi.org/10.1097/PTS.0b013e3182952644

37.   Chuang, H.-H.; Lin, R.-H.; Chen, J.-Y.; Yeh, W.-C.; Lin, H.-F.; Ueng, S.W.-N.; Hsu, K.-H. Effectiveness of a multi-faceted intervention among elementary school children. *Medicine* **2019**, *98*, e15079, https://doi.org/10.1097/MD.0000000000015079.

38.   Kuntz, L.; Scholtes, S.; Sülz, S. Separate & Concentrate: Accounting for Patient Complexity in General Hospitals Supplementary Material. *Manage. Sci.* **2019**, *65*, 2445-2945, https://doi.org/10.1287/mnsc.2018.3064.

39.   Huang, P.-F.; Kung, P.-T.; Chou, W.-Y.; Tsai, W.-C. Characteristics and related factors of emergency department visits, readmission, and hospital transfers of inpatients under a DRG-based payment system: A nationwide cohort study. *PLoS One* **2020**, *15*, e0243373,

https://doi.org/10.1371/journal.pone.0243373.

40.   Verboket, R.; Verboket, C.; Schöffski, O.; Tlatlik, J.; Marzi, I.; Nau, C. Costs and proceeds from patients admitted via the emergency room with mild craniocerebral trauma. *Der Unfallchirurg* **2019**, *122*, 618-625, https://doi.org/10.1007/s00113-018-0566-8.

41.   Jian, W.; Lu, M.; Liu, G.; Chan, K.Y.; Poon, A.N. Beijing's diagnosis-related group payment reform pilot: Impact on quality of acute myocardial infarction care. *Soc. Sci. Med.* **2019**, *243*, 112590, https://doi.org/10.1016/j.socscimed.2019.112590.

42.   Suwei, Y.; Wenwei, L.; Fengqing, W.; Haichen ZHANG, S.W.; Weijun, Z.; Jin, M. Impacts of Hospital Payment Based on Diagnosis Related Groups (DRGs) with Global Budget on Resource Use and Quality of Care: A Case Study in China. *Iran J. Public Health* **2019**, *48*, 238, https://doi.org/10.18502/ijph.v48i2.818

43.   Epstein, A.M. Pay for Performance at the Tipping Point. *N. Engl. J. Med.* **2007**, *356*, 515-517, https://doi.org/10.1056/NEJMe078002.

44.   OECD. Achieving Better Value for Money in Health Care: OECD Health Policy Studies. In Proceedings of Improving Health System Efficiency: Achieving Better Value for Money, Brussels, 17 September; pp. 9-17.

45.   Kortbeek, N.; Zonderland, M.E.; Braaksma, A.; Vliegen, I.M.H.; Boucherie, R.J.; Litvak, N.; Hans, E.W. Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals. *Performance evaluation* **2014**, *80*, 5-26, https://doi.org/10.1016/j.peva.2014.06.003.

46.   Martinussen, P.E.; Hagen, T.P. Reimbursement systems, organisational forms and patient selection: Evidence from day surgery in Norway. *Health Econ Policy Law* **2009**, *4*, 139-158, https://doi.org/10.1017/S1744133109004812

47.   Ellis, R.P. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *J. Health Econ.* **1998**, *17*, 537-555, https://doi.org/10.1016/S0167-6296(97)00042-8

48.   Milcent, C. From downcoding to upcoding: DRG based payment in hospitals. *International Journal of Health Economics and Management* **2021**, *21*, 1-26, https://doi.org/10.1007/s10754-020-09287-x.