

# Role of pay-for-performance in a hospital performance measurement system: a multiple case study in Iran

Aidin Aryankhesal,<sup>1\*</sup> Trevor A Sheldon<sup>2</sup> and Russell Mannion<sup>3</sup>

<sup>1</sup>School of Health Management and Information Sciences, Tehran University of Medical Sciences, Tehran, Iran, <sup>2</sup>Department of Health Sciences, University of York, York, UK and <sup>3</sup>Health Services Management Centre, University of Birmingham, Birmingham, UK

\*Corresponding author. School of Health Management and Information Sciences, Rashid-e Yasemi Ave, Vali-e Asr Ave, Tehran 19956 14111, Iran. Tel: +98-912-4966346. E-mail: a-aryankhesal@tums.ac.ir

---

**Accepted** 5 March 2012

Pay for performance (P4P) is becoming increasingly popular in the health care sector as a tool for encouraging performance (especially quality) improvement. Evidence about the effect of policies in hospitals is rare and generally confined to developed countries. The Iranian hospital grading system, which links the charges hospitals can make for patient stay to the results of their annual performance grading, is one of the earliest examples of P4P in the world. We report here the first evaluation of the impact of the Iranian P4P system.

We conducted a multiple case study using semi-structured interviews and observation in four hospitals with different ownership and grading results, to explore responses to the grading system and the P4P policy. The data were analysed using framework analysis assisted by Atlas-ti software. The findings showed hospital behaviour was influenced by and changed in response to P4P policy, despite serious concerns about the validity of the grading standards. The main driver for such changes was hospital revenue, which acted as a direct financial incentive for private hospital managers and as a factor for public hospital managers' sense of success and reputation. Frontline staff were motivated indirectly by higher revenue flowing into investment in better facilities and working environment. Other potential mechanisms by which the grading system could have influenced behaviour [such as patient and General Practitioner (GP) referral choice] did not appear to influence hospital behaviour.

**Keywords** Pay for performance, performance measurement, hospital behaviour, Iran

---

## KEY MESSAGES

- Pay-for-performance policy resulted in Iranian hospitals increasing their adherence to national measured standards.
- The financial incentives caused hospital managers to make changes even when they were critical of the validity of many of the standards and believed that some were not relevant.
- The financial incentives had similar impact in both public and private hospitals, though for different reasons.

## Introduction

Pay for performance (P4P), or results-based financing (RBF), is increasingly used in some health care systems to improve performance and/or attain targets. The standards used for measuring performance set by payers or the central authorities are usually based on quantity or quality of care, patient experience of care, safety and equity of health care and services, and their cost-effectiveness (Mannion and Davies 2008). Among developed countries, the UK and USA are the pioneers of P4P plans. Since 2004 the UK has used P4P for improving family practitioners' performance (Doran *et al.* 2006). However, in the USA P4P has been more extensively used; more than half of the states had developed P4P plans in their health care systems by 2007 (Kuhmerker and Hartman 2007) and many multidisciplinary medical groups are now paid on this basis (Mullen *et al.* 2009).

USA-based studies have focused on the application of P4P policies in their hospitals (Mehrotra *et al.* 2009) and found that P4P plans were effective in changing hospital behaviour. P4P resulted in increased implementation of some standard guidelines for procedures (Berthiaume *et al.* 2004), employment of new staff and use of new and modern processes (Reiter *et al.* 2006), increase in patients who received appropriate medications (Nahra *et al.* 2006), improved length of stay and patient satisfaction (Berthiaume *et al.* 2006), and improved heart-related outcomes (Grossbart 2006; Lindenauer *et al.* 2007; Sautter *et al.* 2007). Only rarely have studies not found improvements in hospital performance (Glickman *et al.* 2007) or lacked strong evidence on the effectiveness of P4P considerations (Rosenthal *et al.* 2007).

Developing countries have also used P4P plans in their health systems (Loevinsohn and Harding 2005; Eldridge and Palmer 2009; Bredenkamp *et al.* 2011). Reports suggest that this has resulted in some improvements in both quantity or coverage and quality of: (1) institutional or attended delivery in Rwanda, Afghanistan, Cambodia, Haiti and Liberia; (2) antenatal and postnatal care in Rwanda, Cambodia and Afghanistan; (3) immunization coverage, especially for children across Cambodia, Afghanistan and Haiti; (4) family planning programmes in Rwanda, Cambodia and Liberia; (5) preventive care and considerations in Liberia; and also (6) a decrease in family health expenditure through reducing both illegal charges by health centres and the frequency of paying for uneducated traditional care-givers in Cambodia (Soeters and Griffiths 2003; Soeters *et al.* 2006; Eichler *et al.* 2007; Sondorp *et al.* 2009; Toonen *et al.* 2009; Basinga *et al.* 2011; Morgan 2011). All these could result in greater accountability among health care providers and increased technical efficiency, and they could facilitate decentralization in public health systems (Meessen *et al.* 2011). Nevertheless, there are few reported evaluations of the impact of P4P policies on hospital performance in developing countries.

The Iranian health care system introduced P4P in 1997 as part of an annual hospital performance measurement and grading programme. Every year the public medical universities in each province audit all hospitals (public and university owned, private, military, Social Security Organisation or SSO and charity) on behalf of the Ministry of Health and Medical Education (MOHME). Hospitals are graded against standards in 15 domains. These domains include about 1000 questions and

standards, each of which has a defined maximum score. The standards were set at a time when the Iranian hospitals generally lacked necessary equipment and suitable space; so the early focus was naturally on infrastructure rather than other aspects such as quality of procedures. Most measures, even in domains like cleanliness or satisfaction, focused on structure rather than process or outcome to make sure that hospitals addressed the shortage of instruments, equipment and space, although no formal and clear goal was announced when the system was established by the MOHME (MOHME 1997). These standards remain largely unchanged since 1997.

The standards booklets are sent to all hospitals and are also available on the medical universities' websites. The total score achieved plus the separate performance of hospitals in five grading domains together give the grades from +1 to 5, where +1 is the best possible result and 5 is the worst (see Table 1). Hospitals are expected to put a copy of the grading result certificate on all their notice boards—the only formal method stipulated for disseminating grading results. The grade awarded to a hospital determines the amount it can charge for patient stay (hotel charges), but all other charges, such as operations, medication and recognition tests, were independent of the hospital grade when this study was conducted (see Table 2). The public and private hospital stay charges for each grade are announced at the start of each year by the MOHME and the

**Table 1** Domains of the Iranian hospital grading system developed from Aryankhesal and Sheldon (2010)

Domain	% of total score
1 Medical and specialized staff	16.5
2 Following values and religious regulations	9.5
3 Hospital installations and construction	9
4 Emergency department	8.5
5 Medical equipment and medicine	8.5
6 Management	7.5
7 Nursing staff	7.5
8 Other staff	5.5
9 Satisfaction	4.5
10 Medical records and informatics	4.5
11 Hospital committees	4.5
12 Sanitation and cleanliness	4
13 Non-medical equipment	3.5
14 Quality indicators	3.5
15 Safety and security equipment	3
<b>Total</b>	<b>100</b>
16 Teaching activity	2.5
17 Non-general departments	2.5
18 ICU and CCU	2.5
19 Other special facilities	2.5

*Notes:* The shaded domains can affect the total grade of the hospital; if the achieved score in them is less than a specific range, the final grade will fall by a band.

Domains number 16 to 19 are not compulsory, but count as a bonus to hospitals.

ICU: Intensive Care Unit; CCU: Coronary Care Unit.

**Table 2** Financial incentives in the Iranian hospital grading system

Grade	Awarded score (%)	Patient stay charges (10 000 Iranian Rials)*		Explanation
		Public hospitals	Private hospitals	
+1	≥100	14.5	60	No additional financial incentive for this grade over grade 1.
1	85–99	14.5	60	
2	70–84	11.6	48	
3	55–69	8.7	36	
4	50–54	7	24	Hospitals will be graded again after 3 months. If their grade has not improved they will be graded as 5.
5	≤49	The licence of such hospitals will no longer be valid and they can only operate as clinics or minor surgery centres.		

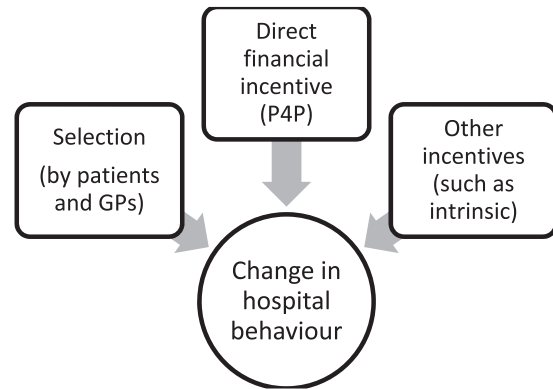
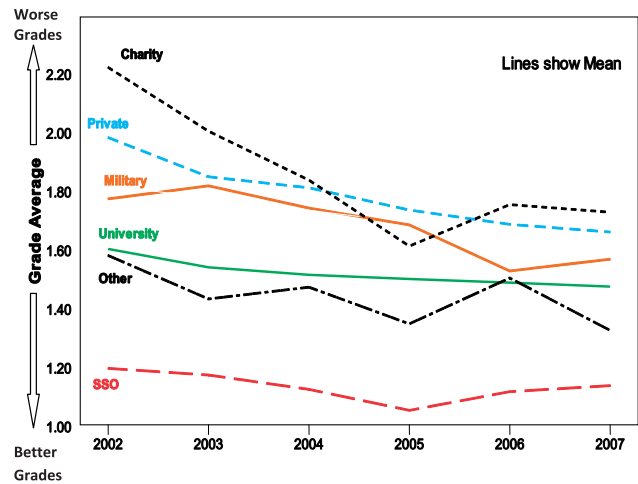
Note: \* Based on 2008 charges for a bed in a room with at least 3 beds.

Iranian Medical Council (which works as a syndicate for physicians).

The proportion of total revenue derived from patient stay charges differs by type of admission. In Iranian hospitals that provide only short stay surgery only 15% of revenue may come from patient stay charges, whilst in intensive care units (ICUs) and chronic care hospitals this may reach 80–85%. The average share of total hospital revenue generated by patient stay charges in a general hospital is about 45% (Medical Services Insurance Company 2010). Therefore, given that patient stay charges increase by 25–33% for a one grade improvement (see Table 2), a one point increase in hospital grade can result in revenue increasing by around 4–28%, depending on the nature of the hospital, with the typical increase being 11–15%, or annually about US\$277 000 to US\$378 000, based on the 2007 average revenue of 24 public hospitals in the studied region (we could not access revenue information of private hospitals).

Hospital stay charges were generally covered by insurance companies; patients paid only up to 10% of stay charges. Therefore, the MOHME's P4P policy was less likely to affect the access of insured poor patients to a highly graded hospital.

The MOHME expects this system to result in hospitals adhering to the grading standards. Figure 1 summarizes the mechanisms through which the grading system could plausibly change hospital behaviour: (1) a selection mechanism in which the grading results influence patients' and General Practitioners' (GPs) choice of hospital; (2) the financial incentive due to the effect on revenue; and (3) other incentives such as intrinsic and personal motives. The national grading results show that hospitals of any ownership type have generally improved their grade over time (Figure 2) (Aryankhesal 2010). This paper reports the first study to examine whether the Iranian P4P policy using the grading system has had an impact on hospital behaviour, and if so through what mechanism. The need for case studies that look at why and how P4P may affect the

**Figure 1** The grading system's potential incentives for hospitals to change their behaviour**Figure 2** Average hospital grade over time by ownership type adapted from Aryankhesal (2010)

behaviour of relevant stakeholders in response to the incentives was indicated by earlier reviewers (Eldridge and Palmer 2009; Mehrotra *et al.* 2009).

Recently we have shown that the grading results alone were unlikely to influence either patient or GP hospital choices, mainly because of very low general awareness of the grading system and hospital grading results (Aryankhesal and Sheldon 2010). Even those small groups of patients (6%) and physicians (12%) who were aware of the grading results did not generally use this information when choosing a hospital, either because factors such as hospital reputation and reputation of a hospital's specialist physicians were more important (to patients), or hospital grading results were not seen as reflecting quality of care or were not trusted (by GPs). If we assume on the basis of this earlier research that changes in hospital behaviour are not driven by this mechanism, hospitals may adopt the grading standards in response to the P4P mechanism and/or the intrinsic and personal values. This study, therefore, examined whether hospital behaviour could possibly be shaped by these two mechanisms (Figure 1).

## Methods

This study was conducted in late 2008 in Tehran province, the capital, which contains about one-fifth of all 800 Iranian hospitals. There are three geographical regions in this province, with the hospitals in each supervised by a medical university; the biggest one (located in the West) with 67 hospitals was selected. We conducted a multiple case study in this region. The goal was to compare the role of P4P policy in public (university) and private hospitals, as well as in hospitals whose grades had improved or fallen at their last grading round. This purposive sampling would let us obtain sharper findings about the role of the MOHME's P4P policy. Therefore our sample would include two public and two private hospitals, one of each with an improved rating and one of each with a fallen rating.

Given the above-mentioned inclusion criteria, we reviewed hospitals' last grading result in the selected region. Among the 26 university hospitals graded by the medical university in Tehran province, there was one hospital with an improved grade (from 2 to 1) in the 2007 grading round and this was selected for study. However, no public hospital grade fell. Consequently, we decided to select one of the public hospitals that had remained in grade 2 for seven consecutive years (the poorest result for a public hospital in the region). Among the 28 private hospitals there were two with improved grades; one from 2 to 1 and one from 3 to 2. The former was selected owing to its greater success. Also, there were only two private hospitals where grades had fallen, both from 1 to 2, one of which refused to participate in the study. Table 3 summarizes characteristics of the included hospitals.

Data were collected mainly through interviews and notes taken as part of a site visit. In each hospital a wide range of staff was interviewed, including the top level (hospital head/CEO, hospital manager), middle (nursing managers, department heads) and frontline staff (physicians, nurses and other staff). There was a high level of participation, with around 90% agreeing to be interviewed—64 staff in total. Owing to the complexity of concepts, we used semi-structured interviews which explored staff awareness of the grading system, asked about their views of the grading system standards and the related incentives for staff, and examined if there were changes triggered by the grading system and their effects on quality of care. Observation notes were taken for two general purposes. First, we wanted to verify the changes mentioned by interviewees as their hospital's response to the grading system. Most of the changes mentioned were structural and physical changes, such as room renovations and modern equipment being installed. One researcher (AA) attended the wards and

departments where the changes were made and took note of the quality of such changes to see if they were the same the interviewees had reported. Secondly, notes were taken in order to describe the physical features of the hospitals, which could give a general picture when combined with other information from interviews.

All interviews were undertaken at the interviewees' working hospital, mostly in a quiet place where only the interviewer and interviewee were present. However, in some cases the interviewees, especially nurses, could not leave their work place and some interruptions in the interviews were inevitable. Interviews were recorded, unless the interviewee did not give consent, in which case notes were taken instead.

All interviewees received an information sheet before the interview that introduced the researcher and outlined the goals of the research, the voluntary nature of participation in the study and the anonymity of participants. Interviewees' informed consent was recorded at the beginning of the interview or was signed if only notes were taken. At the end of each interview the key points mentioned by the interviewee were summarized for him/her to check whether the perceived meanings were the same as what they intended (member checking) (Lincoln and Guba 1985). The average duration of each interview was 41 minutes.

All interviews were transcribed and the observation notes were added to the transcripts. Due to the study's semi-structured topic guide and generally pre-defined codes developed before analysis from the general theoretical framework (Figure 1), framework analysis was performed rather than thematic analysis or grounded theory which are used when interview questions are more open (Pope *et al.* 2006). Atlas-ti 6 was used to facilitate the framework analysis. The quantitative data on awareness of the grading system and results were analysed using a chi-square ( $\chi^2$ ) test in SPSS version 18. All the names of hospitals and participants were coded and anonymity was preserved at all stages.

The researchers had no previous familiarity with any of the interviewees or any association with the studied hospitals. The research study was approved by the Research Governance Committee, Department of Health Sciences, University of York, UK.

## Results

### Awareness

Most staff in each hospital knew their hospital's grade, but awareness of the detailed scores and accessibility of the grading standards booklet was not that high (see Table 4), especially in the university hospitals (A and B), where only the managers

**Table 3** The characteristics of the selected hospitals

Hospital	Ownership	Location	No. of beds	Last year grading change	Year of establishment
A	Public	Tehran Province, a town suburb	130	Improved from 2 to 1 (Dec 2007)	1966
B	Public	Tehran Province, a town	100	No change, grade 2 for 7 consecutive years until Dec 2008	2000
C	Private	Tehran Province, Tehran	100	Improved from 2 to 1 (Jun 2008)	1969
D	Private	Tehran Province, Karaj	305	Dropped from 1 to 2 (May 2008)	2000

**Table 4** Hospital staff awareness of the grading system

	Hospital A (n = 21)	Hospital B (n = 17)	Hospital C (n = 15)*	Hospital D (n = 10)
Awareness of their hospital's grade (%)	95.2	94.1	100	70
Awareness of the awarded score to their unit/ward (%)	19	23.5	40	60
Awareness of accessibility of the grading standards booklet (%)	28.6	29.4	60	50

Note: \* One of the interviewees in this hospital joined a paired interview just after the awareness measurement section and so was excluded from this table.

and some middle level staff knew the detailed scores and the availability of the grading standards booklet. Distribution of awareness of grading scores was significantly different between the private and university hospitals 48% vs 21.1%, ( $P=0.025$ ,  $\chi^2$  test), as was awareness of accessibility of grading standards, 56% to 28.9% ( $P=0.032$ ,  $\chi^2$  test). There was no significant difference in awareness distribution between hospitals with improved grades and those unimproved. One of the interesting findings was that none of the Emergency Department (ED) physicians, who were general practitioners, had seen the grading standards or were interested in knowing, even though ED is one of the most important domains of all in the grading (see Table 1); they were barely aware of the grading results. Many staff thought that the grading standards were kept secret from hospitals.

*"The grading scores and the grading standards are not presented to me. Of course I have not been curious enough to ask about them as well."* (ED manager and physician, Hospital A)

*"The checklists are private and secret. They [grading teams] do not let us see them."* (Laboratory head, Hospital A)

### Perceived validity

Although staff had little awareness about the actual grading standards, most had been notified about the importance of different issues by their managers and had witnessed the actual grading sessions through which the grading auditors had evaluated their department/unit. Staff judgements (mainly criticisms) of the standards are summarized in Table 5. They questioned particularly their content validity. Most staff believed they focused excessively on measures of structure such as hospital building and medical equipment, rather than process and outcome measures. Also they believed that standards should not be the same for all hospitals, but that they should be based on hospitals' local needs and their ownership.

*"If our hospital gets grade 2 [rather than 1], I would not feel guilty, because having a good building and equipment is not what I can do for the hospital."* (Nurse, Hospital C)

*"The grading auditors asked us to buy a basin washing machine, but we were using disposable basins and do not need such machine. However, they insisted on buying the machine and we had to follow the standard. This caused a lot of financial pressure for us."* (Hospital manager, Hospital D)

They believed that the grading standards were out of date and ambiguous so that no one could judge whether the hospital had met the standards. Hence, auditors could be too

**Table 5** Comparison of views on the validity of the grading system

Type of validity	Validity issue	Hospital			
		A	B	C	D
Content	Excessive focus on structure measures	•	•	•	
	Lack of focus on specialist physicians' performance	•		•	•
	Same standards for the public and private hospitals	•	•	•	
	Irrational insistence on specific amount of equipment		•	•	
	Excessive focus on GPs' performance	•		•	
	Lack of sufficient focus on staff welfare and satisfaction		•	•	
	Negligence of local needs in hospital grading standards		•		•
	Lack of focus on nursing quality	•		•	
Face*	Ambiguity of the grading standards and questions	•	•		
	Superficial grading standards		•	•	•
	Old grading standards			•	•
Grading process	Surveyors' superficial grading standards		•		•
Grading results	Easygoing policy, especially for the university hospitals	•	•		•

Note: \*Face validity is defined as 'investigators' subjective assessment of the presentation and relevance of the questionnaire: do the questions appear to be relevant, reasonable, unambiguous and clear?' (Bowling 2009, p. 167).

subjective. Moreover, certain staff in hospitals that failed to improve their grade (B and D) believed that the auditors examined their hospital superficially so that they could easily pass some of the standards. However, such a belief was not expressed in hospitals A and C which had improved their grade, possibly because they did not like to question the validity of their grade.

*"One of the questions in the nursing domain says 'Do the nurses in the different wards co-operate together?' and it carries 20 points. I will give it 20, but the grading team may give it 15. How will you justify that it should not be 15?"* (Matron, Hospital B)

*"So far the grading team have never asked patients' satisfaction directly from them; they always have received the completed forms from us, which may be falsified."* (Head nurse, Hospital D)

In Iran, most public hospitals are owned by medical universities, which also conduct the grading audits. This raised the concern that medical universities were self-assessing, and so were lenient about their own hospitals' shortcomings. Poor grades would result in lower income and so the medical universities, as the owners and financial supporters of these hospitals, would suffer financially. There could also be reputational reasons for medical universities to give their own hospitals more generous grades, and probably to grade private hospitals, as their competitors, more harshly.

*"We and the grading teams are colleagues. The grading is only a self-evaluation. The charges are affected by the grading results, so they like us to get good grades."* (Head nurse, Hospital A)

*"The public hospitals are graded negligently. If they get poor grades, MOHME will incur a loss."* (Laboratory Manager, Hospital D)

**Incentives**

Staff reported different incentives for following the grading standards. In general, top level staff and most middle level staff mentioned hospital revenue as their main driver. All interviewees acknowledged that improving hospital grade would increase hospital revenue materially (owing to the increase in patient stay charges) and any drop in hospital grade would cause significant revenue reductions. Hospital D managers even mentioned that a drop of hospital grade had forced them to lay off some of their staff in order to cope with the lower income. This was acknowledged by other staff as well.

*"We should try; if hospital grade is not good some will be fired or laid off due to the decrease in hospital revenue."* (Medical records staff, Hospital D)

Financial incentives were felt more directly by the high and middle level staff in the private sector, because they were shareholders and so benefited directly. However, in the university (public) hospitals, where hospital managers derive no direct financial benefit from the hospital's revenue, reputation among authorities, respect and honour resulting from the grade and its good financial status and development programmes were mentioned as the motivation for getting a good grade.

*"Retaining grade 1 is very important for us. If the hospital could not get a good grade it may go bankrupt then shut down... We need to improve the hospital by new wards and departments and these all need money which comes from a good grade."* (Hospital head, Hospital A)

The effect of grading on competitive position was mentioned only in the private hospitals. Their managers thought patients were aware of the grading system and that this influenced their choice of hospital. In contrast, the university hospitals did not believe that local people were aware of the grading results or did not take this into account.

*"Currently the biggest motive for the private hospitals is competition. We must compete to survive and the grading results are one of the main instruments for this. Nowadays people's understanding*

*of treatment issues has increased significantly through TV, internet and satellite programmes."* (Hospital manager, Hospital C)

Among the frontline staff the most explicit incentive was the impact on the working place resulting from a better grade. These staff believed that if their hospital improved its grade, the managers would invest in better facilities and instruments for staff, and improve their office and work place. However, they complained about lack of improvement in their income after their hospital had an improved grade.

*"If the grade improves we will have a better work place and feel more secure and relaxed... Also it is possible that staff receive some facilities at their workplace, such as better instruments and clothes, which makes working here easier."* (Radiology technician, Hospital B)

*"We worked hard for reaching this grade, but, but, our income did not change. We are disappointed now and feel less interested in retaining grade 1."* (Medical records unit staff, Hospital C)

**Changes in hospitals**

Table 6 summarizes the changes hospitals had made in order to improve their grade. These changes were mentioned by at least two interviewees in each hospital and most were verified by observing the evidence during data collection. Most of these changes relate to four out of the five grading domains that were given higher weighting by the MOHME (highlighted in Table 1): ED, medical records and informatics, cleanliness and sanitation, and quality indicators. Several are related to equipment and buildings that are themselves emphasized in the standards.

All hospitals had to some extent made changes in order to retain or improve their grade, although some were unsuccessful (B and D). Aside from a few staff at high level, staff generally thought that without the grading system hospitals would not be making the changes, because many were not directly beneficial for the hospital. Some changes just resulted in extra costs for hospitals, as with the basin washing machine quoted earlier for hospital D, but the hospitals had implemented the changes to do well in the grading.

**Table 6** Reported practical changes to obtain better grades

	Hospital			
	A	B	C	D
New equipment and services	•		•	•
Infection control and cleanliness	•		•	•
New Emergency Department building	•	•		
Waiting time decrease in Emergency Department	•	•		
Improved nursing records	•		•	
Improved security	•			
Patient honouring plans	•			
Less delay in sending the insurance companies' bills		•		
Rooming-in care		•		
Railing system in medical records unit		•		
Increased number of staff				•

*"If there were no grading, the waste would not be separated into yellow and black bags. We nurses also would have less focus [on our jobs]. Even the specialists would have less focus."* (Nurse, Hospital C)

*"We needed a ventilator in the 'dark room' because of the chemicals, and had requested one on many occasions, but they [hospital managers] granted our request only after the grading teams' notification."* (Radiology technician, Hospital D)

## Discussion

This research used a multiple case study design. This is a suitable approach when the main research question is 'how' a system works and 'why', where the researcher has no control over actual behavioural events and the focus is on contemporary vs historical events (Yin 2009). The case study approach has been used previously in some international studies to examine the effects of performance measurement activities or systems (Dziuban *et al.* 1994; Rosenthal *et al.* 1998; Chassin 2002; Mannion *et al.* 2005). By including several cases we were able to reach more robust conclusions about the role of P4P mechanisms in the Iranian hospitals. Sixty-four interviews represents quite a large sample for a qualitative study (Pope and Mays 2006).

We found that the grading system had triggered changes in all hospitals studied. Even hospitals that failed to improve their grade in the most recent grading round had made changes in order to obtain a better grade, but they were just insufficient. Indeed there was hardly any difference in the behaviour of those hospitals which had improved their grade and those which did not. Our analysis of grading data showed that the results were improving over the last few years (Figure 2), generally due to investment in infrastructure over time. In our opinion, these changes were unlikely to have happened without the presence of the grading system. Most occurred because of, or were reinforced by, the P4P system. Two sorts of evidence support this view; the staff's views and the type of changes seen in the hospitals.

Staff believed that without the grading system their hospital would not have implemented most of the changes. They indicated that because many changes increased costs for their hospital with no direct financial benefit, hospital managers would not have made these investments in the absence of the grading system. Changes like improvement of medical records and related processes, cleanliness considerations, and even buying unnecessary equipment and instruments, are examples of such changes. Moreover, staff had been critical of the grading standards, and so would be unlikely to adhere to them if there was no P4P associated with the grading system.

The changes involved were generally those which corresponded to four of the five more highly weighted grading standards, relating to the emergency department, cleanliness, medical records, and quality indicators. Further, hospitals' excessive focus on equipment and building features was at least to some extent the result of the grading system's emphasis on structural standards. As shown in Table 1, the grading instrument included four separate domains related to the structural features: 'hospital installations and construction', 'medical equipment and medicine', 'non-medical equipment' and 'safety and security

equipment', accruing about 9, 8.5, 3.5 and 3% of the total score, respectively, altogether about 24%. However, other grading domains also included some questions related to equipment and physical facilities in different areas of the hospital, which added to the weighting of this factor.

In addition to demonstrating that the grading system influenced hospital behaviour, the study examined the drivers for behaviour change. Three potential drivers were plausible (see Figure 1). Our findings indicated that P4P seems to have the main role. The two other potential drivers—the selection mechanism affecting patient and GP referral choice, and other incentives such as intrinsic and personal motives—were seldom reported. Only in the private hospitals did some of managers feel that patients may be aware of the hospital grading results and that this would influence their choice of hospital accordingly. This would reflect the fact that both private hospitals were in cities where there was some competition. Because the validity of the grading standards was questioned, it is unlikely that many would be motivated to achieve them for intrinsic satisfaction or profound pride. The frontline staff were more motivated by the benefits that additional revenue could bring, such as a better working environment because the hospital could buy better equipment and facilities.

We conclude that the P4P mechanism, resulting in a financial incentive arising from higher patient stay charges for better grades, was the most plausible mechanism in triggering and shaping changes in the studied hospitals. The selection mechanism had been shown as ineffective in our previous research (Aryankhesal and Sheldon 2010) and the intrinsic incentives proved to be less important. The P4P mechanism motivated hospital managers and some middle level staff in both public and private hospitals, but there was no direct financial incentive for the frontline staff. This did not prevent change in the hospitals, probably because of the Iranian hospitals' centralized managerial system. Even if frontline staff were not so motivated directly to follow the grading standards, the top level staff were, so most of the target changes were implemented.

This study was cross-sectional and we examined hospitals just one year after their grading change, so the findings about staff incentives, especially for frontline staff, may not be generalizable to the longer term. If hospitals reach a good grade and can retain it for some time (as is true for several hospitals), the P4P incentive may work differently. This needs further study and comparison of the impact of hospital staff incentives over time.

We examined different hospitals in terms of their ownership, grading results and their location; this is one of the strengths of the study. However, one of the limitations is that we could not include a control hospital where the grading system was run without the P4P mechanism, or where grading was not mandatory. From 1998 all Iranian hospitals have had to undergo the grading evaluation and all have had to follow the charges announced by the authorities. So we cannot say with complete confidence what would have happened in the absence of this policy. Another weakness is that both university (public) hospitals were in rural areas or small towns, whilst both private hospitals were in cities. This may have affected our findings due to differences in local people's socio-economic characteristics. Indeed, hospital managers' beliefs about the influence of the grading results on patients' choice of hospital, and so the

differences in their thinking about public expectations, could possibly be due to location, not ownership. Extending the research to university hospitals in big cities would resolve this issue. However, the grading results in all hospitals and medical universities are broadly similar so it is likely that in the other cities and provinces the findings would be similar.

Established in 1997, the Iranian grading system is probably one of the oldest running examples of 'pay for performance' being used in hospitals. The impact we found is consistent with the results of evaluations in US hospitals (Mehrotra *et al.* 2009). Nearly all these studies showed that P4P plans had positive effects on hospitals' performance (Berthiaume *et al.* 2004; Berthiaume *et al.* 2006; Grossbart 2006; Nahra *et al.* 2006; Reiter *et al.* 2006; Lindenauer *et al.* 2007; Sautter *et al.* 2007). Our findings have some similarities with those from studies examining the impact of P4P policies in primary care in developing countries. These studies found that P4P encouraged target behaviour in delivering organizations (Soeters and Griffiths 2003; Soeters *et al.* 2006; Eichler *et al.* 2007; Sondorp *et al.* 2009; Toonen *et al.* 2009; Basinga *et al.* 2011; Meessen *et al.* 2011; Morgan 2011).

The US-based studies which examined the effect of P4P policies in the context of public reporting of performance data were similar to this study. However, they were either principally quantitative and did not have designs which could assess causality (Berthiaume *et al.* 2004; Berthiaume *et al.* 2006; Grossbart 2006; Nahra *et al.* 2006; Glickman *et al.* 2007; Lindenauer *et al.* 2007), or they had a very small number of interviews focused on a few issues (Reiter *et al.* 2006; Sautter *et al.* 2007), and so could not thoroughly explore whether improvements were a result of the P4P programme. In this study we knew there had been grading 'improvements' over time and we set out to explore why change occurred, in particular the impact of P4P on change.

This study has a number of implications for policy in Iran as well as for other countries implementing or seeking to implement P4P schemes. Most importantly, linking measured performance to hospital prices so that high performing hospitals (as defined by the scheme) can increase revenue can be a powerful mechanism to change the behaviour of hospital managers, and through them the behaviour of the staff. This effect operates even where there is a degree of scepticism by staff as to the validity of the dimensions of performance measured and/or the objectivity of the method of performance measurement. Direct incentives to health care organizations to improve performance do not need to be accompanied by the publication of hospital performance to have an impact and so may be useful where patients and other stakeholders do not have choice or are unlikely to use performance data in making hospital choices.

What is not clear from this or most other studies of P4P is the cost-effectiveness of such a mechanism, in other words whether the improvement in performance stimulated by the incentive mechanism justifies the extra expenditure due to increased tariffs in better performing organizations. This will depend to an extent on what dimensions of performance are included in the scheme and how well they are measured. If the grading standards are not valid (i.e. are not likely to reflect good quality care) and the process for measuring performance is not fair and

efficient then the value of P4P schemes will be limited. The Iranian example demonstrates how important it is to constantly generate and reflect on emerging evidence, and to revise standards to ensure that the best outcomes in terms of quality of care are likely to be achieved given the cost constraints. Whilst the strong emphasis on infrastructure (such as equipment) might have been a valid priority at the time the grading policy was introduced, it now results in over-investment in this area and insufficient emphasis on effective clinical processes and outcomes.

This research is the first to study the role of P4P mechanisms in changing hospital behaviour in Iran, and unusually for studies of P4P in developing countries, it explores the qualitative aspects in order to better understand why and how the system works. We conclude that the P4P system changed hospital behaviour, but whether this results in improvement of the quality of care is out of our study's scope. Further research on the impact of the Iranian grading system and its P4P mechanism on clinical indicators would be useful.

## Funding

Aidin Aryankhesal was supported by Iranian Ministry of Health and Medical Education and Tehran University of Medical Sciences.

## Conflict of interest

None declared.

## References

- Aryankhesal A. 2010. The Iranian hospital grading system and its influence on stakeholders' behaviour. PhD thesis, University of York, UK.
- Aryankhesal A, Sheldon T. 2010. Effect of the Iranian hospital grading system on patients' and general practitioners' behaviour: an examination of awareness, belief and choice. *Health Services Management Research* **23**: 139–44.
- Basinga P, Gertler P, Binagwaho A *et al.* 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet* **377**: 1421–28.
- Berthiaume J, Chung R, Ryskina K, Walsh J, Legorreta A. 2006. Aligning financial incentives with quality of care in the hospital setting. *Journal of Healthcare Quality* **28**: 36–44, 51.
- Berthiaume J, Tyler P, Ng-Osorio J, Labresh K. 2004. Aligning financial incentives with Get with the Guidelines to improve cardiovascular care. *American Journal of Managed Care* **10**: 501–4.
- Bowling A. 2009. *Research Methods in Health*. Maidenhead, UK: Open University Press.
- Bredenkamp C, Borman ND, Mullen P *et al.* 2011. Dealing with difficult design decisions: The experience of an RBF pilot program in Haut-Katanga District of Democratic Republic of Congo (DRC). The World Bank Group. Online at: <http://www.rbhealth.org/rbhealth/news/item/500/dealing-difficult-design-decisions-rbf-pilot-program-democratic-republic-congo-drc>, accessed 31 July 2011.
- Chassin MR. 2002. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Affairs* **21**: 40–51.



- Doran T, Gravelle H, Reeves D *et al.* 2006. Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine* **355**: 375–84.
- Dziuban JSW, McIlduff JB, Miller SJ, Dal RH. 1994. How a New York cardiac surgery program uses outcomes data. *The Annals of Thoracic Surgery* **58**: 1871–6.
- Eichler R, Auxila P, Antoine U, Desmangles B. 2007. Performance-based incentives for health: six years of results from supply-side programs in Haiti. CGD Working Paper #121. Washington, DC: Center for Global Development.
- Eldridge C, Palmer N. 2009. Performance-based payment: some reflections on the discourse, evidence and unanswered questions. *Health Policy and Planning* **24**: 160–6.
- Glickman S, Ou F, Delong E *et al.* 2007. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *Journal of the American Medical Association* **297**: 2373–80.
- Grossbart S. 2006. What's the return? Assessing the effect of "pay-for-performance" initiatives on the quality of care delivery. *Medical Care Research and Review* **63**(Suppl. 1):295–485.
- Kuhmerker K, Hartman T. 2007. Pay-for-performance in state Medicaid programs: a survey of state Medicaid directors and programs. New York: The Commonwealth Fund, April 2007.
- Lincoln YS, Guba EG. 1985. *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications.
- Lindenauer P, Remus D, Roman S *et al.* 2007. Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine* **356**: 486–96.
- Loevinsohn B, Harding A. 2005. Buying results? Contracting for health service delivery in developing countries. *The Lancet* **366**: 676–81.
- Mannion R, Davies H. 2008. Payment for performance in health care. *British Medical Journal* **7639**: 306–8.
- Mannion R, Davies H, Marshall M. 2005. Impact of star performance ratings in English acute hospital trusts. *Journal of Health Services Research & Policy* **10**: 18–24.
- Medical Services Insurance Company. 2010. Applied process of grading hospitals and Intensive Care Units. Tehran: Medical Services Insurance Company, Office of Health Services Evaluation.
- Meessen B, Soucat A, Sekabaraga C. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bulletin of the World Health Organization* **89**: 153–6.
- Mehrotra A, Damberg C, Sorbero M, Teleki S. 2009. Pay for performance in the hospital setting: what is the state of the evidence? *American Journal of Medical Quality* **24**: 19–28.
- MOHME. 1997. *The instruction of standards and principles of evaluation of the general hospitals*. Tehran: The Iranian Ministry of Health and Medical Education.
- Morgan L. 2011. When the hustle gets rough: making performance-based contracting work in Liberia. The World Bank Group. Online at: <http://www.rbflhealth.org/rbflhealth/news/item/513/when-hustle-gets-rough-making-performance-based-contracting-work-liberia>, accessed 31 July 2011.
- Mullen K, Frank R, Rosenthal M. 2009. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *NBER Working Paper Series*. Cambridge: National Bureau of Economic Research.
- Nahra T, Reiter K, Hirth R, Shermer J, Wheeler J. 2006. Cost-effectiveness of hospital pay-for-performance incentives. *Medical Care Research and Review* **63**(Suppl. 1):49S–72S.
- Pope C, Mays N. 2006. *Qualitative Research in Health Care*. London: BMJ Books.
- Pope C, Ziebland S, Mays N. 2006. Analysing qualitative data. In: Pope C, Mays N (eds). *Qualitative Research in Health Care*. London: BMJ Books.
- Reiter K, Nahra T, Alexander J, Wheeler J. 2006. Hospital responses to pay-for-performance incentives. *Health Services Management Research* **19**: 123–34.
- Rosenthal GE, Hammar PJ, Way LE *et al.* 1998. Using hospital performance data in quality improvement: the Cleveland Health Quality Choice experience. *The Joint Commission Journal on Quality Improvement* **24**: 347–60.
- Rosenthal MB, Landon BE, Howitt K, Song HSR, Epstein AM. 2007. Climbing up the pay-for-performance learning curve: where are the early adopters now? *Health Affairs* **26**: 1674–82.
- Sautter K, Bokhour B, White B *et al.* 2007. The early experience of a hospital-based pay-for-performance program. *Journal of Healthcare Management* **52**: 95–107.
- Soeters R, Griffiths F. 2003. Improving government health services through contract management: a case from Cambodia. *Health Policy and Planning* **18**: 74–83.
- Soeters R, Habineza C, Peerenboom PB. 2006. Performance-based financing and changing the district health system: experience from Rwanda. *Bulletin of the World Health Organization* **84**: 884–9.
- Sondorp E, Palmer N, Strong L, Wali A. 2009. Afghanistan: Paying NGOs for performance in a postconflict setting. In: Eichler R, Levine R & The Performance-Based Incentives Working Group (eds). *Performance Incentives for Global Health: Potential and Pitfalls*. Washington, DC: Center for Global Development.
- Toonen J, Canavan A, Vergeer P, Elovainio R. 2009. *Performance-based Financing for Health: Lessons from Sub-Saharan Africa*. Amsterdam: Royal Tropical Institute, Cordaid and the World Health Organization.
- Yin RK. 2009. *Case Study Research Design and Methods*. Thousand Oaks, CA: Sage Publications.