Leadership in Health Services

A systemic approach to quality improvement in public health services: The Moroccan "Quality Contest"

Amina Sahel Abdelali Alaoui Belghiti Vincent DeBrouwere Filomena Valente Soares Guy Kegels Nejoua Belkaab Isabelle Godin Sabine Ablefoni Anselm Schneider Bruno Dujardin

Received 27 July 2013
Revised 13 November 2013
Accepted 11 April 2014

A systemic approach to quality improvement in public health services

The Moroccan "Quality Contest"

Amina Sahel, Abdelali Alaoui Belghiti, Vincent DeBrouwere, Filomena Valente Soares, Guy Kegels, Nejoua Belkaab, Isabelle Godin, Sabine Ablefoni, Anselm Schneider and Bruno Dujardin
(Author affiliations can be found at the end of the article)

Abstract

Purpose – The purpose of this paper is to discuss the results of the first four years of implementation of a quality program called "Quality Contest" (QC). This program was implemented from 2007 onward to improve the quality of hospital services by the Moroccan Ministry of Health. The peculiarity of this intervention, held every 18 months, is that it combines several approaches (self-evaluation, external audits with feedback, hospital ranking, awards and performance disclosure) and focuses on the quality of management.

Design/methodology/approach – The assessment tool used to evaluate the quality of hospital management consists of 80 items. In each contest, a score is attributed to each item based on the score given for self-evaluation and the score given by external auditors. The sum of these scores allows the global performance score of the hospital to be obtained. To compare the performances over time and among hospitals, Wilcoxon signed-rank, Wilcoxon–Mann–Whitney and Kruskal–Wallis statistical tests were used.

Findings – The results of the QC organized between 2007 and 2010 revealed that the hospitals participating in all the three contests had significantly improved their performance levels in terms of the quality of management. There was also a significant association between the number of times hospitals participated in the QC and the performance scores attained.

Originality/value – The paper reports an original quality improvement approach in a developing country that succeeded in triggering sustainable improvement dynamics by combining support (feedback) with reward (prizes) and pressure measures (ranking, performance disclosure).

Keywords Hospitals, Quality management, Organizational performance

Paper type Research paper

Leadership in Health Services
Vol. 28 No. 1, 2015
pp. 8-23

1751-1879
DOI 10.1108/LHS-07-2013-0033

Introduction

Concern about the healthcare quality and healthcare services performance has been increasing among policymakers and insurers, as well as health managers. In response to this concern, a growing number of quality models and approaches have been developed over the past few decades (Lanier *et al.*, 2003). Despite having the same goal, the interventions and principles underlying these quality approaches are rather varied (Groene *et al.* 2008; Ovretveit and Gustafson, 2003). Quality systems may target the individual provider's performance or the entire organization's performance. They may

focus on the procedure or on the results of care, utilize self-evaluation or external auditing, provide feedback privately or publish results and, lastly, favor financial (pay for performance) or symbolic (prizes, certificates, etc.) incentives.

In Morocco, pilot experiments for quality improvement were first implemented in the healthcare services in the 1990s, with several approaches being tested (i.e. quality circles, team-based problem-solving, clinical audits) (Blaise, 2005; Muffler *et al.*, 2007).

Although these projects brought about improvements, these were often observed to wane upon finalization of the project. Internal evaluation identified several reasons for the lack of sustainability, namely, the absence of recognition and valorization of the work of the teams involved in these approaches, and the fact that managers were not accountable for the progress of quality improvement actions. Furthermore, these experiments only focused on either input or output improvement, never on the management process, and only involved one part of the health system, either hospitals or primary health centers. To remedy these inadequacies, the Moroccan Ministry of Health (MOH) introduced in 2007, on a national scale, a quality improvement program called the "Quality Contest" (QC), inspired by the systemic quality improvement approach developed by the German Technical Cooperation Agency (Schneider and Stierle, 2007). The first interest of the QC is to focus on the quality of the management process (organization and functioning), which is an important component that must be taken into account; otherwise, all investments (equipment, staffing) that Ministries of Health of developing countries may make would barely have an effect on the quality of services. The hypothesis was that by improving the process, the results would also improve. The second interest of the QC is its combination of several approaches: self-evaluation, external audits, personalized feedback, hospital ranking with awards attributed to the best, support (training and supervision) given to the weakest and results published in a report distributed to all participating hospitals. The two components, hospital ranking and disclosure of results, were considered by the MOH as leverages for eliciting professional accountability and governance, which are often lacking in the healthcare systems of developing countries.

This article aims to present the results of a nation-wide intervention combining complementary approaches and discuss the underlying mechanisms that contributed to produce its effects.

Methods

Description of the intervention

The QC comprises seven steps (Figure 1). The first step consists of informing all hospitals about the QC and its rules. The second step is the "self-evaluation", aiming at actively involving the participating teams in the process and encouraging them to discuss and reflect on their practices. In this step, the teams are called upon to explain and describe in a document, called the self-evaluation document, and their management practices according to each item. This document is then sent to the central evaluation unit where analysts score each item on the basis of only what was reported in the document (third step).

Following this phase, the audit and feedback phase (fourth step) is initiated with teams of external auditors visiting the participating hospitals to evaluate their management process and provide a second score for the audited items. As the audit visit lasts just one day, only a sample of items are audited and scored. The audit is concluded

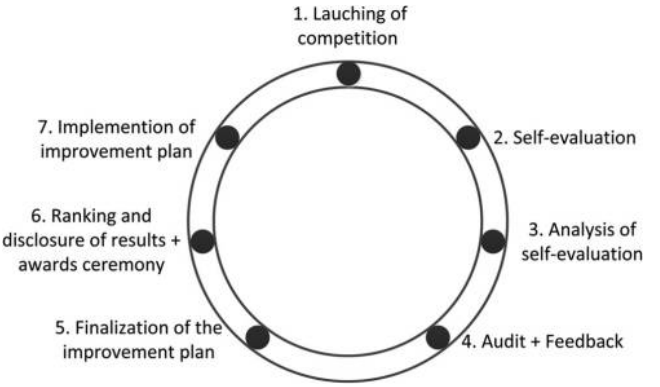

Figure 1.
Steps of the QC in Morocco

in a meeting with the local team, so as to present a feedback on the strengths of the hospital and its weaknesses requiring correction. On this basis, an improvement plan is jointly drafted. This draft is later finalized by the participating hospitals (5th step) and acts as a guide for the necessary improvements. Based on the scores obtained by each hospital, a ranking is established, and a ceremony presided over by the Minister and attended by representatives from other departments and international organizations takes place at the end of each contest to announce the results and distribute awards to the best-performing hospitals (6th step). Following each contest, the results are published in a detailed report that is then distributed to the participating hospitals, thus making the results openly accessible. The final step is the implementation of the "improvement plan" by the participating hospitals.

At first, the QC was based on the voluntary participation, but from the third contest onward, it became compulsory. Between 2007 and 2010, three contests were organized nationally (2007, 2008 and 2010). The results of the fourth QC are not yet available.

Developing the quality scores

To develop the assessment tool, professionals and field managers were involved in an interactive procedure that took one year, so as to define the domains to be evaluated, considered important in the management quality of their services. The domains identified were "client satisfaction", "ethics", "accessibility/availability/continuity of services", "rationalization of resources", "safety/reactivity", "continuous improvement", "skills development" and "leadership". Each of the selected domains was subdivided in two to four aspects or subdomains, describing the level of quality to be achieved (Table I).

Each aspect was then divided into several items, which were formulated as questions according to the management steps of the Deming wheel: plan, do, check, act (Deming, 1986). Finally, about 80 items were retained.

For each item/question, a range of expected responses was defined to facilitate scoring. This range of expectations indicated the desired level of quality on a scale from 0 to 4 (0: no action; 1: minimal level; 4: the maximum desired level). In each contest, an initial score (analysis score) between 0 and 4 was assigned to each item by the analysts based on the self-evaluation document submitted. A second score (audit score) was then given by the external auditors, who evaluated the management process on the site for a

Domains	Subdomains
Domain 1: Client satisfaction	A system of listening to the complaints and expectations of clients is introduced
	A mechanism for staff motivation is implemented
	The internal environment of the hospital is comfortable
Domain 2: Ethics	A mechanism to assure the respect of ethics is implemented and is functional
	The missions of social assistance are properly carried out
Domain 3: Accessibility/availability/continuity	The access to the hospital departments is organized and facilitated
	The mechanisms to improve the availability of resources are established and functional
	The referral system is established and functional
	The regulation of emergencies is effective
Domain 4: Rationalization of resources	The control tools, in accordance with the MOH policy orientations (hospital reform, drugs policy, etc.), exist and are functional
	The hospital applies a strategy of maintenance of the real estate
	The hospital applies mechanisms to reduce the waste of resources
	The hospital ensures the monitoring and assessment of outsourcing contracts
Domain 5: Safety/reactivity	A management system of hospital waste is put in place and is functional
	A mechanism of prevention and control of nosocomial infections is put in place
	A mechanism for reduction of occupational risks is put in place
	A mechanism to ensure patient safety exists and is functional
Domain 6: Continuous improvement	The quality improvement process is put in place
	Operational research exists and is encouraged
Domain 7: Skills development	A strategy for human resources management is defined and operational
	The existing guidelines of best practices are disseminated, known and applied
	The mechanisms for the development of standard operating procedures exist and are functional
Domain 8: Leadership	The process of establishment of business plan is engaged
	The management bodies are established and exercise their functions
	The hospital has a communication plan
	4. The hospital initiates partnership relations

Table I.
Domains and subdomains evaluated in the QC

sample of items through observation, discussion and reading of available materials. The analysis and audit scores per item for each hospital were entered into a database.

Because only a sample of items per hospital has been audited, the other items do not have an audit score and, therefore, a final score. To overcome this constraint, it was considered that if all the items had been audited in a hospital, the average difference between the audit and analysis scores would substantially be the same as that obtained for the sample of audited items in this hospital. This average difference was used as a correction factor (Fc) and applied to the analysis scores of this hospital:

$$Fc = \frac{\sum \Delta i}{\text{NB of audited items}}$$

Where $0\,i = S_{audit} - S_{analysis}$ (the difference between the audit and the analysis score for each audited item).

The Fc was then added to the analysis scores to obtain a final score per item called the corrected final score (CFSi).

$CFSi = S_{analysis} + Fc$.

The sum of all the *CFS*i is the global score (GS) of the hospital: $GS = \sum CFSi$

To facilitate comparisons among hospitals and across years, the GS was expressed as the performance level attained:

$$\textit{Performance level attained} \text{ (per cent)} = \frac{\text{GS X 100}}{\text{Total no. of items X 4}}$$

where (4) is the maximum score per item, and the (total number of items $\times$ 4) is the maximum attainable performance score for a hospital.

When hospitals inflate their self-assessment, their analysis scores will be higher than their audit scores, and therefore, their Fc will be negative. When they underestimate their performance, their analysis scores will be lower than the audit scores, and their Fc will be positive. As the nature of the contest provides incentive for hospitals to obtain high performance, it would be easy for some of them to inflate their self-assessment. However, this risk is controlled by the application of the correction factor. Indeed, a hospital that has inflated its self-assessment, would get a negative Fc and would have its GS corrected and lowered.

Table II show how the GS is calculated with an example of a hospital that inflated its self-assessment. To simplify the calculation, we consider, in this example, that the total number of items is ten and the sample of items audited is four.

Analysis

An initial descriptive analysis (medians and 95 per cent confidence interval [CI]) allowed us to describe the performance level attained by the hospital and its year.

We made two aggregated comparisons. The first is based on the number of times hospitals have participated in the contest. The hypothesis was that hospitals which participated three times should have better scores than those that participated two times, and the latter should have better scores than those that participated only once.

To verify this hypothesis, we compared the 2008 performance scores of hospitals that participated once with those of hospitals which participated twice. We used the

Evaluated items	Analysis score ($S_{analysis}$)	Audit score (S_{audit})	$\Delta i = S_{audit} - S_{analysis}$	$CFS_i = S_{analysis} + F_c$**
Item 1	3	2	–1	3 + (–1) = 2
Item 2	2			2 + (–1) = 1
Item 3	2			2 + (–1) = 1
Item 4	1			1 + (–1) = 0
Item 5	4	3	–1	4 + (–1) = 3
Item 6	0			0 + (–1) = –1
Item 7	4			4 + (–1) = 3
Item 8	2	0	–2	2 + (–1) = 1
Item 9	0			0 + (–1) = –1
Item 10	1	1	0	1 + (–1) = 0
Total for the hospital	19		–4*	9***

Notes: *$\Sigma\Delta$i i.e.: (–1) + (–1) + (–2) + (0) = –4; **$FC = \Sigma\Delta i / NB\ of\ audited\ items$ i.e.: (− 1) + (− 1) + (− 2) + (0) / 4 = − 1; ***ΣCFS_i (GS for hospital) i.e.: 2 + 1 + 1 + 0 + 3 – 1 + 3 + 1 – 1 + 0 = 9; The sum of scores assigned to self-assessment is 19. After correction, the GS for hospital is 9. Expressed as performance level attained, this GS = 22.5%: Performance level attained (per cent) = GS × 100 / Total no. of items × 4 = 9 × 100 / 10 × 4 = 22.5 per cent

Table II. Method to calculate the GC (with an example of hospital that inflated its self-assessment)

Mann–Whitney–Wilcoxon test that allows a comparison of a continuous variable (the 2008 performances scores) between two unrelated groups, in this case, hospitals that participated once and hospitals that participated twice. We also compared the scores achieved in 2010 by hospitals that participated once with those of hospitals which participated two and those which participated three times. To do so, we used the Kruskal–Wallis test which allows comparison of a continuous variable (the 2010 performance scores) among three or more unrelated groups (in this case, hospitals with a single participation, hospitals with two participations and hospitals with three participations). To see whether this trend is linear, we used the Cuzik's trend test.

The second comparison is based on the evolution of scores between 2007 and 2010 for hospitals that have participated in the three QCs. We hypothesize that the repeated and regular participation in the QC enables hospitals to improve their performance over time. To verify this hypothesis, we used the Wilcoxon signed-rank test which allows for the comparison of paired data, in this case, performance scores of hospitals that have participated in the three QCs, between 2007 and 2010.

Finally, to control a possible selection bias (hospitals which participated in all QC and improved their performance were the best at the beginning), we compared the 2007 performance of these hospitals with the performance of those which have not been regular (i.e. only participated in either one QC or two QCs). To do this, we used the Mann–Withney–Wilcoxon test which allows the comparison of a continuous variable (performance in 2007) between two unrelated groups (in this case, hospitals which are regular and those which are not).

We also made two disaggregated comparisons by domain.

First, we used Wilcoxon signed-rank test to compare the performance by domain attained by hospitals which participated in the three QCs between 2007 and 2010.

Second, we used the Kruskal–Wallis test to compare the performance by domain achieved in 2010 according to the number of participations in the QC.

Results

Global performance

Among the 96 regional and provincial public hospitals in Morocco, 47 hospitals participated in the 2007 QC (first QC), 53 in 2008 (second QC) and 94 in 2010 (third QC).

If such an approach is effective, an improvement in the performance level should be observed in the hospitals most often participating in the QC.

Thus, for the second contest (2008), we compared the performance of hospitals participating for the first time ($n = 14$) with those participating for the second time ($n = 39$). These results showed a difference in the performance between the former (39 per cent; 95 per cent CI: 36–50 per cent) and the latter (50 per cent; 95 per cent CI: 43–56 per cent), but this difference was not significant ($p = 0.102$; Mann–Whitney–Wilcoxon test).

A comparison of performances attained by hospitals in the 2010 contest according to the number of participations (Figure 2) revealed a significant difference in performance ($p < 0.001$; Kruskal–Wallis test) among the hospitals ($N = 39$) participating for the first time (35 per cent; 95 per cent CI: 33-40 per cent), those ($N = 18$) participating for the second time (42 per cent; 95 per cent CI: 39-49 per cent) and those ($N = 37$) participating for the third time (57 per cent; 95 per cent CI: 52-63 per cent), with this association being linear ($p = 0.000$; Cuzik's trend test).

The weak performance of the 39 hospitals participating for the first time in the 2010 contest may have been influenced by the fact that this group largely comprised small-sized (rural) hospitals, a possible confounding factor. However, when excluding small-sized hospitals, we were able to confirm the significant difference ($p = 0.003$) in the performance among the hospitals participating once ($N = 15$, 42 per cent; 95 per cent CI: 34-50 per cent), twice ($N = 15$, 43 per cent; 95 per cent CI: 41-54 per cent) and thrice ($N = 36$, 56 per cent; 95 per cent CI: 51-63 per cent). This same analysis could not be applied to small-sized hospitals because of their low numbers in 2007 and 2008.

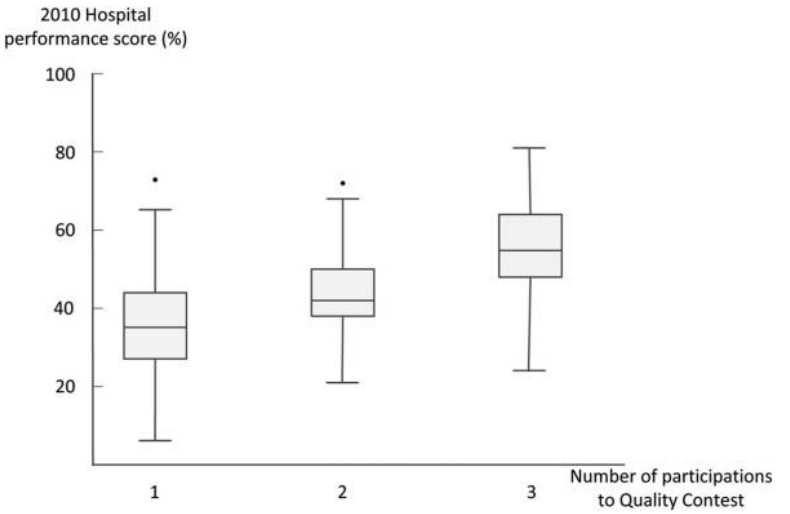

Figure 2.
Hospital performance scores obtained in 2010, according to the number of participations

The result also revealed that the performance attained by the 37 hospitals (Figure 3) participating regularly (in the three QCs), significantly improved between 2007 and 2010 ($p = 0.038$; Wilcoxon signed-rank test) with a narrowing of variance: 47 per cent in 2007 (95 per cent CI: 39-58 per cent) and 57 per cent in 2010 (95 per cent CI: 52–63 per cent).

To verify whether the hospitals continuing to participate regularly in the QC were initially better than those participating irregularly (selection bias), the 2007 performance of the former ($N = 37$; 47 per cent; 95 per cent CI: 39-58 per cent) was compared with that of the latter ($N = 10$; 44 per cent; 95 per cent CI: 22-62 per cent). However, the results did not show any significant difference ($p = 0.39$; Mann–Whitney–Wilcoxon test) which indicate that there is no selection bias.

These statistical tests are summarized in Table III.

Performance by domain

We then investigated which domains have improved significantly between 2007 and 2010. We found that in hospitals participating regularly ($n = 37$), every domain improved over time (Table IV). However, this improvement was significant (Wilcoxon signed-rank test) only for four domains:

(1) Domain 4, "rationalization of resources" ($p = 0.047$);

(2) Domain 5, "safety/reactivity" ($p = 0.015$);

(3) Domain 7, "skills development" ($p = 0.002$); and

(4) Domain 8, "leadership" ($p = 0.013$).

It is not significant for "client satisfaction", "ethics", "accessibility/availability/continuity of services" and "continuous improvement".

We also compared the scores achieved in the 2010 contest according to the number of participations for each domain. The results showed a significant difference in performance scores ($p < 0.001$) among the hospitals participating for the first time, those participating for the second time and those participating for the third time (Table V);

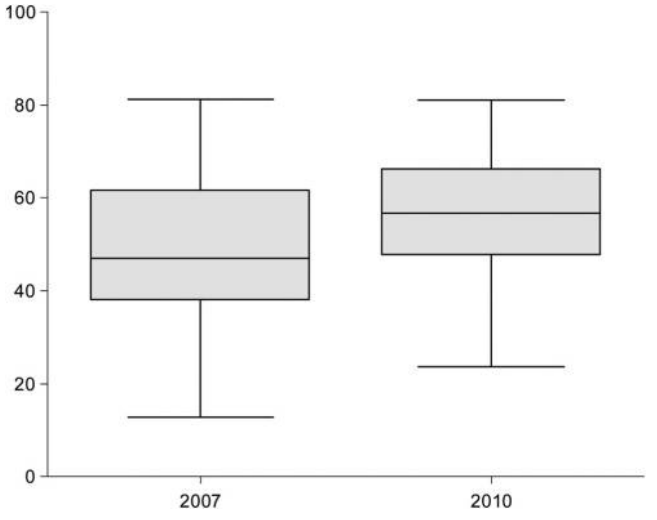

Figure 3.
Evolution of performance scores between 2007 and 2010 for hospitals participating in the three QCs ($N = 37$)

Table III.
Summary of statistical tests

Responding variable	Null hypothesis	Statistical conclusion	Tests	*p*-value (level of significance = 0.05)
2008 performance scores	Scores distribution in the hospitals participating for the first time = scores distribution in the hospitals participating for the second time	H_0 not rejected (there is no difference in the scores distribution between the two groups)	Mann–Whitney–Wilcoxon	0.102
2010 performance scores	The hospitals participating for the first time, those participating for the second time and those participating for the third time have the same scores distribution	H_0 rejected (there is a difference in performance scores among the three groups)	Kruskal–Wallis	< 0.001
Performance scores of the hospital participating in the 3 QC	Scores distribution in 2007 = scores distribution in 2010	H_0 rejected (for hospitals participating in the three QCs, there is a difference between their 2009 and 2010 performance)	Wilcoxon signed-rank	0.038
2007 Performance scores	The scores distribution for the hospitals continuing to participate regularly = scores distribution for the hospitals participating irregularly	H_0 not rejected (there was no difference between the two groups at the start)	Wilcoxon–Mann–Whitney	0.39

with this association being linear ($p = 0.000$), being involved in the QC apparently enables hospitals to improve their performance at each participation.

Discussion

The result by domain showed that scores improved significantly over time except for four domains (1, 2, 3 and 6). The non-significant increase of the scores in these domains between 2007 and 2010 may reflect the unexpected difficulties met by the hospital managers in implementing some actions related to these domains. These difficulties may be financial. This may be the case for Domain 1: the improvement of hospital environment needs additional budget not easy to mobilize by hospitals. Constraints may also be related to the difficulty to coordinate with external actors. This may be the case for Domain 3: the referral system and the regional organization of emergencies do not depend solely on the hospital. Domain 2 is related to the respect of ethical principles by the health personnel, and such a change in personnel behavior may take time, probably more than a couple of years; this may explain the absence of significant improvement. Finally, the time needed to install a research culture and develop research skills may explain the absence of improvement for Domain 6.

However, even if the improvement is not significant for these four Domains between 2007 and 2010 in the group of hospitals which have regularly participated in the QC, differences in these domains are, nevertheless, significant when comparing these hospitals with those that have participated only once or twice in the QC. This suggests that regular participation in the QC may make the difference, although the statistical significance would appear only after a longer period of time.

For the other domains and for the global performance, the improvement is significant. Indeed, our results suggest that participating in the QC allowed hospitals to improve their performance scores, with this improvement being linked with the number of participations. There did not appear to be a selection bias for hospitals at the outset, as those continuing to participate had no significantly different initial performances compared to the others which have stopped participating to the QC.

To explain the observed improvements, several hypotheses may be put forth. These hypotheses arise from the different QC components (i.e. audit and feedback, follow-up of the improvement plans, competition and ranking, publication of the results and prize distribution), which were each designed and integrated into the intervention to stimulate the quality improvement of services.

The first of these components was the so-called "self-evaluation", whereby the participating hospital's team was required to reflect on their practices regarding each

	2007	2010	*p*-value (Wilcoxon signed-rank test)
Domain 1	53% (95% CI: 44–58%)	60% (95% CI: 54–66%)	0.531
Domain 2	50% (95% CI: 38–56%)	53% (95% CI: 45–60%)	0.274
Domain 3	46% (95% CI: 41–57%)	55% (95% CI: 50–59%)	0.158
Domain 4	51% (95% CI: 41–61%)	59% (95% CI: 53–63%)	0.047
Domain 5	49% (95% CI: 37–53%)	62% (95% CI: 52–70%)	0.015
Domain 6	49% (95% CI: 45–56%)	51% (95% CI: 44–64%)	0.723
Domain 7	48% (95% CI: 42–59%)	64% (95% CI: 52–68%)	0.002
Domain 8	46% (95% CI: 37–52%)	58% (95% CI: 53–66%)	0.013

Table IV. Evolution by domain of performance scores for hospitals participating in three QC ($N = 37$)

Table V.
Performance scores by domain obtained in the 2010 contest according to the number of participations in the QC

Domains	Median performance and 95% CI obtained in 2010 by the hospitals participating for the first time ($N = 39$)	Median performance and 95% CI obtained in 2010 by the hospitals participating for the second time ($N = 18$)	Median performance and 95% CI obtained in 2010 by the hospitals participating for the third time ($N = 37$)	p-value (Kruskal–Wallis test)
Domain 1	39% (95% CI: 34–44%)	52% (95% CI: 45–59%)	60% (95% CI: 54–66%)	< 0.001
Domain 2	31% (95% CI: 26–35%)	37% (95% CI: 31–49%)	53% (95% CI: 45–60%)	
Domain 3	32% (95% CI: 28–45%)	42% (95% CI: 34–49%)	55% (95% CI: 51–59%)	
Domain 4	36% (95% CI: 28–45%)	47% (95% CI: 34–54%)	59% (95% CI: 53–63%)	
Domain 5	40% (95% CI: 31–49%)	46% (95% CI: 39–51%)	62% (95% CI: 52-70%)	
Domain 6	30% (95% CI: 22–34%)	34% (95% CI: 24–47%)	51% (95% CI: 44–64%)	
Domain 7	42% (95% CI: 35–47%)	50% (95% CI: 40–60%)	64% (95% CI: 52–68%)	
Domain 8	34% (95% CI: 30–39%)	45% (95% CI: 38–52%)	58% (95% CI: 53–66%)	

item and explain and discuss them before describing them in the self-evaluation document. This process may have constituted the first step toward questioning local practices and moving toward change.

The second component of the QC that may have contributed toward improvements was the feedback given at the end of audit to participating hospitals, concerning their strengths and weaknesses. The aim of this feedback was to make the teams aware of their weaknesses, thus allowing them to act and bring about the necessary changes. Although the effects of auditing and feedback are generally considered to be weak to moderate (Flottorp *et al.*, 2010), a well-conducted feedback has been recognized to be a fundamental element for improving practices (Bradley *et al.*, 2004; Jencks *et al.*, 2003). A qualitative study (Bradley *et al.*, 2004) conducted among professionals and managers identified some conditions necessary for feedback to promote change, notably, the perception by professionals regarding the validity of the information communicated, the credibility of those identifying problems and giving feedback and the manner of presenting the feedback (supportive or punitive, aimed at the person or system).

The third component of the QC that may explain the observed improvements concerns the hospital's drafting of an improvement plan based on the weaknesses identified in each contest and its subsequent implementation. It appears that actions determined on the basis of problems with objective results made visible to the staff have more chance of being implemented (Kotagal *et al.*, 2009).

The dissemination of the results as part of the QC may have also contributed to hospitals making efforts toward improvements. However, in Western countries where the publication of performance's results is increasingly a part of initiatives to improve quality (Hamblin, 2007;Werner and Asch, 2005), the impact of performance disclosure on changes (in professional behavior, efforts for improvement and outcomes) appears rather mixed, with some studies showing positive results and others showing no effect (Fung *et al.*, 2008;Hibbard *et al.*, 2003,2005;Ketelaar *et al.*, 2011;Marshall *et al.*, 2000;Tu *et al.*, 2009).

The ranking and prize distribution may also have contributed to improving performance. Recognizing and valuing efforts were reported to be motivating factors affecting performance sustainability among healthcare professionals (Dieleman *et al.*, 2006). Similarly, comparing performances among hospitals appeared to be associated with an improvement in care procedures (Merle *et al.*, 2009). Moreover, competition between professionals was described as an incentive toward improvement, either because professionals wish to appear better than their peers or in situations where patients had the option of choosing among providers (Hamblin, 2007). In the same way, participation in contests with the aim of winning a prize may represent an incentive for professionals to improve their procedures and services (Milakovich, 2004).

We did not find an intervention similar to ours that encompasses all these approaches in the literature. However, if the impact of these approaches when implemented separately seems to be quite positive on improving quality, we can expect that their combination in the same intervention would increase the impact, especially in maintaining longer-term dynamics observed in hospitals across the country.

Finally, we think that the bottom-up approach in selecting the items used for the assessment contributed to the dynamics of the improvement observed in hospitals. Indeed, the domains and subdomains proposed by the healthcare professionals

themselves and retained in consensus with the decision-makers made the approach acceptable and contributed to its sustainability.

Although this study showed that hospitals regularly participating in the QC significantly improved their performance scores, we do not know what motivated them to participate in each contest, while others did not. Was it an unfavorable context? The pressure of the contest and a fear of failure? Or a lack of confidence in the objectivity of the approach and in the validity and reliability of the assessment tool? How did the teams experience this procedure and perceive the approach? These are open questions that deserve to be further explored in the framework of a future study. Another issue deserving particular attention is how the observed improvements could have been obtained at lower cost. The QC requires considerable resources, meaning that it may be difficult to sustain in the long term.

Study limitations

The main limitation of our study was the fact that hospital participation was initially voluntary. It became mandatory from the third contest when the first results indicated some improvements.

When the MOH decided to implement the QC, its primary concern was to rapidly improve the quality of care and not to contribute to the scientific evidence. As a consequence, the QC implementation design has not been conceived as a research project, and there was no baseline as such. The baseline consisted of the first evaluation (first QC in 2007) which was an acceptable proxy of the facility performance "before" the improvement plan was implemented. Theoretically, a cluster randomized controlled trial (RCT) would have been the best design to demonstrate the effectiveness of the QC on quality improvement. However, deciding that a random selection of control districts should not benefit from a quality improvement intervention was not politically and ethically acceptable. Such a position is not unusual in policymaking, and the recent debate about the Avahan HIV prevention project implemented in a non-randomized way showed that it is possible – and acceptable – not to use an RCT design to evaluate the effectiveness of a nationwide prevention intervention (Laga and Moodie 2012). RCT designs are the golden standard for evidence-based medicine but are questionable for large public health interventions when the context is an essential component of success (Kemm, 2006;Neuhauser and Diaz, 2007;Victora *et al.*, 2004).

Moreover, the aim of our study was to evaluate whether implementing a complex intervention (i.e. QC) which combines support with reward and pressure measures promoted change. For this reason, our paper did not examine the validity of the assessment tool used. Although this may be considered a limitation, the assessment tool was developed in collaboration with on-site professionals based on their definitions of the key areas of quality needing evaluation. This appears to enhance the relevance and acceptability of this tool, which are features as important as validity in an approach aimed at improvement.

Another limitation is that during the last decade, the Moroccan MOH has implemented a series of reforms aiming at improving the governance of services (education, health, justice, etc.) and at developing more client-centered services. It is possible that the observed results in the QC are partially linked to the overall trend in the society toward quality improvement of services. It is quite difficult to attribute the cause of improvement specifically to this global development or to the QC.

Conclusion

The results of this study suggest that the QC organized in Morocco promoted quality improvement in the management process of hospitals. The combination of approaches involving support (feedback, training), rewards (prices) and persuasive measures (hospitals ranking and performance disclosure) has probably been crucial in creating and maintaining the dynamic of improvement in healthcare services.

The first results of the QC trigger further research questions. For example, which mechanisms really played a role in the observed changes? How did these mechanisms operate? How did they interact with context elements to produce the expected effect or not? Why did these mechanisms work in the context of certain hospitals and not in others?

Such information would be of great interest to avoid unthinking and frustrating replication of "successful" experiments with little chance of succeeding, such as when a particular process, which is possible only in a given context yet essential to the outcome, is not feasible in the situations where the ("initially successful") experiment is to be repeated space between mathematical operators.

References

Blaise, P. (2005), "Culture qualité et organisation bureaucratique, le défi du changement dans les systèmes publics de santé. Une évaluation réaliste de projets qualité en Afrique", *Thèse science de la santé publique*, ESP-Université Libre de Bruxelles, Bruxelles.

Bradley, E.H., Holmboe, E.S., Mattera, J.A., Roumanis, S.A., Radford, M.J. and Krumholz, H.M. (2004), "Data feedback efforts in quality improvement: lessons learned from US hospitals", *Quality and Safety in Health Care*, Vol. 13 No 1, pp. 26-31.

Deming, W.E. (1986), "Out of the crisis: quality, productivity and competitive position", Cambridge University Press, Cambridge.

Dieleman, M., Toonen, J., Toure, H. and Martineau, T. (2006), "The match between motivation and performance management of health sector workers in Mali", *Human Resource for Health*, Vol. 4, No. 2, pp. 23-27.

Flottorp, S.A., Gibis, B., Jamtvedt, G. and McKee, M. (2010), *Utilisation de l'audit et du feed-back aux professionnels de la santé pour améliorer la qualité et la sécurité des soins de santé*, Organisation Mondiale de la Santé and Organisation mondiale de la Santé pour l'Observatoire européen des systèmes et des politiques de santé, Copenhague.

Fung, C.H., Lim, Y.W., Mattke, S., Damberg, C. and Shekelle, P.G. (2008), "Systematic review: the evidence that publishing patient care performance data improves quality of care", *Annals of Internal Medicine*, Vol. 148 No. 2, pp. 111-123.

Groene, O., Skau, J.K. and Frolich, A. (2008), "An international review of projects on hospital performance assessment", *International Journal of Health Care Quality*, Vol. 20 No. 3, pp. 162-171.

Hamblin, R. (2007), "Publishing 'quality' measures: how it works and when it does not?", *International Journal of Health Care Quality*, Vol. 19 No. 4, pp. 183-186.

Hibbard, J.H., Stockard, J. and Tusler, M. (2003), "Does publicizing hospital performance stimulate quality improvement efforts?", *Health Affairs (Millwood)*, Vol. 22 No. 2, pp. 84-94.

Hibbard, J.H., Stockard, J. and Tusler, M. (2005), "Hospital performance reports: impact on quality, market share, and reputation", *Health Affairs (Millwood)*, Vol. 24 No. 4, pp. 1150-1160.

Jencks, S.F., Huff, E.D. and Cuerdon, T. (2003), Change in the quality of care delivered to medicare beneficiaries, 1998-1999 to 2000-2001, *JAMA: The Journal of the American Medical Association*, Vol. 289 No. 3, pp. 305-312.

Kemm, J. (2006), "The limitations of 'evidence-based' public health", *Journal of Evaluation in Clinic Practice*, Vol. 12 No. 3, pp. 319-324.

Ketelaar, N.A., Faber, M.J., Flottorp, S., Rygh, L.H., Deane, K.H. and Eccles, M.P. (2011), "Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations", *The Cochrane Database of Systematic Reviews*, Vol. 11, CD004538.

Kotagal, M., Lee, P., Habiyakare, C., Dusabe, R., Kanama, P., Epino, H.M., Rich, M.L. and Farmer, P.E. (2009), "Improving quality in resource poor settings: observational study from rural Rwanda", *British Medical Journal*, Vol. 339, p. b3488.

Laga, M. and Moodie, R. (2012), "Avahan and impact assessment", *Lancet*, Vol. 379 No. 9820, pp. 1003-1004.

Lanier, D.C., Roland, M., Burstin, H. and Knottnerus, J.A. (2003), "Doctor performance and public accountability", *Lancet*, Vol. 362 No. 9393, pp. 1404-1408.

Marshall, M.N., Shekelle, P.G., Leatherman, S. and Brook, R.H. (2000), "The public release of performance data: what do we expect to gain? A review of the evidence", *JAMA*, Vol. 283, No. 14, pp. 1866-1874.

Merle, V., Moret, L., Pidhorz, L., Dujardin, F., Gouin, F., Josset, V., Graveleau, S., Petit, J., Riou, F., Lombrail, P. and Czernichow, P. (2009), "Does comparison of performance lead to better care? A pilot observational study in patients admitted for hip fracture in three French public hospitals", *International Journal of Health Care Quality*, Vol. 21 No. 5, pp. 321-329.

Milakovich, M. (2004), "Rewarding quality and innovation: awards, charters, and international standards as catalysts for change", in Wimmer, M. Ed., *In Knowledge Management in Electronic Government*, Springer, Berlin, pp. 80-90.

Muffler, N., Trabelssi, M.H. and De Brouwere, V. (2007), "Scaling up clinical audits of obstetric cases in Morocco", *Tropical Medicine & International Health*, Vol. 12 No. 10, pp. 1248-1257.

Neuhauser, D. and Diaz, M. (2007), "Quality improvement research: are randomised trials necessary?", *Quality & Safety in Health Care*, Vol. 16 No. 1, pp. 77-80.

Ovretveit, J. and Gustafson, D. (2003), "Using research to inform quality programmes", *BMJ*, Vol. 326 No. 7392, pp. 759-761.

Schneider, A. and Stierle, F. (2007), *How to Initiate and Steer Systemic Quality Improvement: An Advisor's Guide for the Health Sector and other Social Sectors*, GTZ, Eschborn.

Tu, J.V., Donovan, L.R., Lee, D.S., Wang, J.T., Austin, P.C., Alter, D.A. and Ko, D.T. (2009), "Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial", *JAMA*, Vol. 302 No. 21, pp. 2330-2337.

Victora, C.G., Habicht, J.P. and Bryce, J. (2004), "Evidence-based public health: moving beyond randomized trials", *American Journal of Public Health*, Vol. 94 No. 3, pp. 400-405.

Werner, R.M. and Asch, D.A. (2005), The unintended consequences of publicly reporting quality information. *JAMA*, 293, (10) 1239-1244.

Author affiliations

Amina Sahel, School of Public Health, Université Libre de Bruxelles, Brussels, Belgium and Direction des Hôpitaux et des Soins Ambulatoires, Ministry of Health, Rabat, Morocco

Abdelali Alaoui Belghiti, Direction des Hôpitaux et des Soins Ambulatoires, Ministry of Health, Rabat, Morocco

Vincent DeBrouwere, Department of Public Health, Institute of Tropical Medicine, Antwerpen, Belgium
Filomena Valente Soares, School of Public Health, Université Libre de Bruxelles, Brussels, Belgium
Guy Kegels, Department of Public Health, Institute of Tropical Medicine, Antwerpen, Belgium
Nejoua Belkaab, Direction des Hôpitaux et des Soins Ambulatoires, Ministry of Health, Rabat, Morocco
Isabelle Godin, School of Public Health, Université Libre de Bruxelles, Brussels, Belgium
Sabine Ablefoni, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Eschborn, Germany
Anselm Schneider, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Bonn, Germany, and
Bruno Dujardin, School of Public Health, Université Libre de Bruxelles, Brussels, Belgium

Corresponding author
Amina Sahel can be contacted at: aminasahel@gmail.com