Effectiveness of internal quality assurance programmes in improving clinical practice and reducing costs

Juan J. Gascón Cánovas PhD,1 Pedro J. Saturno Hernández PhD1 and Jose J. Antón Botella PhD2

1Professor, 2Associated Professor, Department of Public Health, School of Medicine, University of Murcia, Murcia, Spain

Abstract

Rationale, aims and objectives To assess the effectiveness of internal initiatives to improve quality as compared with external feedback, and as compared with a control group.

Method Ten primary health centres were randomly selected from the centres in the Murcia’s Region and were randomly assigned into three groups: G1 committed themselves to the improvement; G2 composed by two subgroups: with and without quality improvement (QI) activities – received external feedback; G3 received no intervention. Quality of common cold management was measured in a random sample of 50 patients per centre before and after interventions. Effect was assessed comparing criteria compliance and the cost of treatments for common cold in the three groups.

Results G1 and G2 subgroup with internal QI improved significantly in all criteria, and in G1 average treatment cost decreased by 60% (P < 0.01). G3 improved only in one criterion. Estimated yearly savings in treatment costs for common cold, for a centre with internal QI, were €12 193.83 and it would be €1 817 004.65 for the 30 centres in the Region.

Conclusions Internally assumed QI activities were significantly more effective than external feedback. Besides, the high cost of deficient quality provides a wide margin to invest with benefits in the promotion of internal QI programmes.

Introduction

Quality of care may be assessed to establish what type of actions can be implemented to improve. A systematic review of interventions to improve professional practice showed that effectiveness of audit and feedback ranged from nil to moderate [1–3]. It seems to be that these interventions are not inherently effective and whether they succeed depends on the circumstances in which they are used and how will these interventions themselves be perceived: will quality assurance (QA) programmes be viewed as a ‘hassle’ or a threat, or will they be seen as an opportunity for improvement [4]?

Assessment for improvement can be focused from an internal perspective (internal QA programmes) where the providers of health care are the ones who take on the initiative and responsibility of the activities, or from an external approach (external QA programmes), where groups or institutions not directly involved with the care assessed, usually higher levels to the organization evaluated, are the ones in charge of formulating the rules and doing or making sure the assessment is done [5].

However, the effectiveness in QA depends on the positive change of behaviours and ways of organizing health care delivery [6]. In this sense, it has been hypothesized that this change is favoured when those professionals who provide care are involved in its assessment and improvement [7]. Consequently, it would be logical to state that quality improvement especially depends on programmes with an internal approach [8].

The overview of effectiveness’ interventions for improving quality of care suggests that multifaceted interventions (combination that includes audit or feedback with local consensus process or participation by doctors) are more likely to change clinical practice [2]. But, few studies compared the relative effectiveness of these different strategies [9]. As a result, there is still little scientific evidence on the comparative effectiveness of these methods and their diverse alternatives in practice [10,11].

In this study, we measure the effectiveness of QA activities in centres that undertook them internally, as compared with others in which intervention consisted of an external assessment and feedback of their clinical performance, and as compared with a control group. Effectiveness measurements include compliance with quality of care criteria, rational use of drugs and cost of treatments in cases of common cold – an opportunity to improve selected internally in one of the groups of centres.

Methods

In the framework of a wider study designed to assess, among other issues, the effectiveness of different methods of training for QA, 10 health centres, a simple random sample of the existing 30 in the...
health care system of the Region of Murcia (1 197 646 inhabitants) at the moment of the study, are assigned to three groups by a simple random sampling:

1. **Group 1 (G1). Centres that received training on QA methods and selected the opportunity to improve (internal quality assurance programme group).** Over 1-week period, doctors of these intervention group attended to a course on quality improvement methods [5]. During the training seminar the doctors of these centres selected the common cold management as the opportunity to improve using the suggested guidelines of the Joint Commission on Accreditation of Health Care Organizations (high prevalence and/or high risk for the health and/or frequent tendency to errors). They committed themselves to perform a complete cycle of assessment and improvement. Implementation of the full QA cycle included: (i) problem identification; (ii) study design; (iii) intervention design, implementation of remedial actions; and (iv) re-evaluation. Initially, three centres were assigned to this group, although eventually we only used the data from one of them in the study. One of the centres showed significant improvement, but it was excluded from the study because its information system could not identify the cases to evaluate retrospectively, and they collected data for their improvement cycle in a prospective way designing a specific case identification system for the purpose of the study, that was therefore not comparable to the rest who did it in a retrospective way. In another centre the existence of the opportunity to improve on the chosen issue was dismissed, because the first assessment showed that quality was good, which led to their improvement cycle on another issue.

2. **Group 2 (G2). Centres that received the results of external assessment without their having selected the opportunity to improve (external quality assurance programme group).** In this intervention, clinicians did not participate in the QA programme, they only received individualized information, in tables and graphs, on included performance of the quality criteria in their centre and in the other centres evaluated, although the later were not identifiable. Five centres were assigned to this group, 3 with and 2 without training in QA methods, constituting actually two distinct subgroups. In all of them it was possible to do external assessments with the same methodology.

3. **Group 3 (G3). Control group.** It was formed by two centres, without training in QA and without QA programme. Doctors did not know the topic to be assessed and improve. It was possible to obtain comparable data in both of them.

Any aspect of the research project was unknown by providers of G2 and G3. Doctors of G1 did not have any notice about the existence of G2 and G3 and only knew the objective of the method used for themselves, but did not know the aim of this search. The methodology used for quality assessment followed the model proposed by Palmer [5]. The components of the evaluation plan for all the centres are the following:

---

**Table 1 Quality of care criteria for common cold**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
<th>Exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 1</td>
<td>Physical examination (ear, throat, nose and chest) should be performed.</td>
<td>–</td>
</tr>
<tr>
<td>Criterion 2</td>
<td>Prescribed treatment should be recorded.</td>
<td>–</td>
</tr>
<tr>
<td>Criterion 3</td>
<td>Antibiotics, systemic antihistaminics, and corticoids should not be prescribed.</td>
<td>Co-morbidity: cardiopathy, diabetes, chronic pulmonary disease, immunodeficiency.</td>
</tr>
</tbody>
</table>

---

The units of study were first visits of patients with a diagnosis of common cold (upper respiratory infection), including also those with erythematous pharyngitis and non-stridulous laryngitis. The professionals evaluated were the general practitioners of each health centre. Cases were extracted from the lists of patients who visited the centres in the month previous to the evaluation, and diagnosed with upper respiratory infection.

The cases to be assessed were identified by selecting randomly a sample of 50 cases per health centre proportionally stratified by doctors.

The criteria used to assess quality of care (Table 1) were defined by practitioners belonging to G1, who were the only ones who also chose, through the Nominal Group Technique [12], the care for common cold cases as the opportunity to improve.

Data collected included compliance with the quality of care criteria. Additionally (not envisaged in the criteria of the QA group), the name and characteristics of the prescribed medicines were included in the external evaluations to calculate the cost of pharmaceutical treatments.

The data source used was the medical record of the cases patient selected in the sample. A total of 401 and 396 medical records were evaluated in the first assessment and in the reassessment respectively. The price of drugs prescribed was obtained from the International Handbook of Pharmaceutical Drugs.

Two set of measurements were made: a first evaluation and then a second evaluation, 6 months after implementing interventions (internal discussion following their internal initiative in G1, feedback of quality criteria compliance in G2). G1 was doubly measured, first internally and checked then externally by the same evaluator who audited the rest of the centres. All data were collected by one training research. In this external evaluation the name and characteristics of the prescriptions were noted down to calculate treatment cost, but these data were neither used in internal assessment nor included in the feedback.

To assess the criterion of the appropriateness of the pharmacological treatment, cases were excluded in which the treatment had not been recorded or where there was some accepted exception to the criterion (see Table 1). The total cost of treatments was calculated summing up the prices of the medicines prescribed.

To measure the effectiveness of the interventions, we used the difference in the proportion of criteria compliance (second evaluation–first evaluation) and also the differences in treatment cost/care process, within every study group, between the first and second evaluation. Differences in criteria compliance were calculated in absolute terms (the difference between both assessments) and in relative terms as the relative percentage of improvement, defined as the absolute improvement in relation to the amount of potential improvement at the outset. The later is calculated
substracting from 100% the percentage of performance in the first evaluation. The relative percentage of improvement takes into account the base line and expresses comparably the effect of the improvement effort carried out.

The cost of deficient quality was estimated as the average cost of inadequate drugs treatments prescribed (which include antibiotics, histaminic mono drugs or corticoids), multiplied by the proportion of inadequate treatments and the number of annual cases of common cold recorded. It has been calculated for each group of centres and on a regional level for the year of the study (726,788 cases recorded by the difference of average pharmaceutical cost (in cases of acute respiratory infection of ear, nose, throat and chest) for improvement (which include antibiotics, histaminic mono drugs or corticoids), multiplied by the proportion of inadequate treatments and the number of annual cases of common cold recorded. It has been calculated for each group of centres and on a regional level for the year of the study (726,788 cases of acute respiratory infection of ear, nose, throat and chest recorded in the Region). The minimum potential savings were calculated in each study group multiplying the number of annual cases recorded by the difference of average pharmaceutical cost (in the second assessment), between each one of the study groups and the group with QA activities, which was considered as reference.

The precision in the estimates has been calculated with a 95% confidence. For the statistical analysis of the differences between the two evaluations we used the value of \( z \) in the comparison of two proportions and two means [13], to assess the improvement in the performance of the quality criteria evaluated and the decrease of the cost of the therapeutical process, respectively. We use one-tail tests given that the alternative hypothesis for the interventions, in this case the existence of improvement, is unidirectional. Tests were performed using SPSS/PC v. 15.0.

### Results

#### First evaluation: variability in quality of care between the centres and the existence of room for improvement

Compliance with the assessed quality of care criteria by centre in the first evaluation ranged between 23.6% and 95.0%. There are marked, but not homogeneous, differences between the three groups: the centre with QA activities starts with higher levels of performance than those of the other centres for the criterion of physical examination (Table 2), while having the lowest level in the criterion of treatment recording (Table 3), and a medium value as regards to treatment adequacy (Table 4) and its cost (Table 5). On the whole, despite the differences between the groups for each criterion individually considered, in all of them the criterion less compiled is the physical examination, and the one that is most fulfilled is the recording of the treatment. In all the groups there was a wide margin for improvement in relation to the diagnostic

### Table 2 Performance of physical examination. % compliance by group of centres

<table>
<thead>
<tr>
<th></th>
<th>First evaluation (1)</th>
<th>Second evaluation (2)</th>
<th>( \Delta (2) - (1) )</th>
<th>Relative improvement*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
</tr>
<tr>
<td>Group 1 (internal QA) (c = 1; ( n_c = 51; n_r = 50 ))</td>
<td>52.9 ± 13.7</td>
<td>72.0 ± 12.4</td>
<td>19.1 ± 18.6</td>
<td>40.6 ± 39.5</td>
</tr>
<tr>
<td>Group 2 (feedback)   (c = 5; ( n_c = 250; n_r = 246 ))</td>
<td>23.6 ± 5.3</td>
<td>32.5 ± 5.9</td>
<td>9.9 ± 7.9</td>
<td>11.6 ± 10.3</td>
</tr>
<tr>
<td>Group 2.1 (centres with training in QA methods) (c = 3; ( n_c = 150; n_r = 146 ))</td>
<td>22.0 ± 6.6</td>
<td>43.1 ± 8.0</td>
<td>21.1 ± 10.4</td>
<td>27.1 ± 13.4</td>
</tr>
<tr>
<td>Group 2.2 (centres without training in QA methods) (c = 2; ( n_c = 100; n_r = 100 ))</td>
<td>26.0 ± 8.6</td>
<td>17.0 ± 7.4</td>
<td>−9.0 ± 11.3</td>
<td>−12.2 ± 15.3</td>
</tr>
<tr>
<td>Group 3 (control) (c = 2; ( n_c = 100; n_r = 100 ))</td>
<td>40.0 ± 9.6</td>
<td>56.0 ± 9.7</td>
<td>16.0 ± 13.7</td>
<td>26.7 ± 22.8</td>
</tr>
<tr>
<td>Average (c = 8; ( n_c = 401; n_r = 396 ))</td>
<td>31.4 ± 4.5</td>
<td>43.4 ± 4.9</td>
<td>12.0 ± 6.0</td>
<td>17.5 ± 8.7</td>
</tr>
</tbody>
</table>

* \( \Delta = \frac{n_r - n_c}{n_c} \)

### Table 3 Treatment recorded. % compliance by group of centres

<table>
<thead>
<tr>
<th></th>
<th>First evaluation (1)</th>
<th>Second evaluation (2)</th>
<th>( \Delta (2) - (1) )</th>
<th>Relative improvement*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
<td>(±95% CI)</td>
</tr>
<tr>
<td>Group 1 (internal QA) (c = 1; ( n_c = 51; n_r = 50 ))</td>
<td>72.5 ± 12.3</td>
<td>86.0 ± 9.6</td>
<td>13.5 ± 15.6</td>
<td>49.1 ± 56.7</td>
</tr>
<tr>
<td>Group 2 (feedback)   (c = 5; ( n_c = 250; n_r = 246 ))</td>
<td>93.2 ± 3.1</td>
<td>94.3 ± 2.9</td>
<td>1.1 ± 4.3</td>
<td>16.2 ± 63.2</td>
</tr>
<tr>
<td>Group 2.1 Centres with training in QA methods (c = 3; ( n_c = 150; n_r = 146 ))</td>
<td>93.3 ± 4.0</td>
<td>97.9 ± 2.3</td>
<td>4.6 ± 4.6</td>
<td>69.2 ± 69.1</td>
</tr>
<tr>
<td>Group 2.2 (centres without training in QA methods) (c = 2; ( n_c = 100; n_r = 100 ))</td>
<td>93.0 ± 5.0</td>
<td>89.0 ± 6.0</td>
<td>−4.0 ± 7.0</td>
<td>−57.1 ± 90.0</td>
</tr>
<tr>
<td>Group 3 (control) (c = 2; ( n_c = 100; n_r = 100 ))</td>
<td>95.0 ± 4.3</td>
<td>94.0 ± 4.0</td>
<td>−1.0 ± 6.3</td>
<td>−20.0 ± 26.0</td>
</tr>
<tr>
<td>Average (c = 8; ( n_c = 401; n_r = 396 ))</td>
<td>91.0 ± 2.8</td>
<td>93.2 ± 2.5</td>
<td>2.2 ± 3.7</td>
<td>24.4 ± 41.1</td>
</tr>
</tbody>
</table>

* \( \Delta = \frac{n_r - n_c}{n_c} \)

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd
Second evaluation: differences in criteria compliance after the intervention

The centre with internal QA activities underwent a significant improvement in the three evaluated criteria: physical examination (Table 2) increased by 19.1 points (from 52.9 ± 13.7 to 72.0 ± 12.4; \( P = 0.02 \)); treatment recording (Table 3) increased by 13.5 points (from 72.5 ± 12.3% to 86.0 ± 9.6%; \( P = 0.05 \)); and treatment adequacy (Table 4) improved by 22.6 ± 17.6 points (from 67.6 ± 15.1% to 90.2 ± 9.1; \( P = 0.04 \)).

In the feedback and control groups we observe an increase only in the compliance with the criterion related to physical examination (Table 2). In the feedback group it increased 8.9 points (from 40.0 ± 9.6% to 56.0 ± 9.7%; \( P = 0.01 \)). Although all the groups experienced a significant improvement in performing physical examination, it is in the internal QA group where there is a higher relative improvement (40.6%) as compared with the control group (26.7%) and the feedback group (11.6%). The centres of the feedback group with training in QA methods also improved significantly in the other two criteria, while the feedback centres without QA training responded in a way similar to the control groups.

Differences in treatment cost after the intervention

In the internal QA centre there is a significant decrease in the cost/care process of €1,85 (\( P < 0.01 \)), while there were no significant changes neither in the feedback group nor in the control group. The estimated annual savings in the centre that took on internally QA activities amount to €12,193.83 (Table 5). If all the health centres of the Region should reach the level of treatment quality achieved by the centre with internal QA activities after the improvement (second evaluation), the potential annual savings in the treatment of the common cold, in the Region, would be €1,817,004.65 (Table 6).

Discussion

Comparability of data and validity of results

Although the evaluation criteria, except for the cost assessment of treatments, were defined by G1 and they were the ones who evaluated their own cases, the comparability was assured by doing the same assessment externally in all the centres with the same criteria and by the same evaluator. In the G1 centre the assessment was double (internal and external) without the results being different; the assessment of treatment cost was carried out externally in all cases and without the centres having any knowledge of this fact. On the other hand, to further assure comparability, we only used...
the data from one centre of the three foreseen in G1, because it was not possible to use identical methodology to evaluate in the other two. This resulted in a decrease in the number of G1 cases, which can cast doubt about the grade of generalization of the results; however, the randomness of the assignment of the centres to the various groups, the net distinction between them in relation to the intervention carried out and the consistency of the results (reinforced by what was obtained in the subgroup of centres of G2 with training and previous activities in QA) suggest that the behaviour of the G1 centre adequately represents the expected result with a similar intervention in centres with the same characteristics. Another factor that can affect comparability between the various centres, and for that reason, generalization of the data, is the variability itself in the level of performance of the criteria at the baseline (first assessment). In this sense, the situation would be unfavourable for G1 as, when assessing less cases, significant differences may be hindered in their appearing (these have to be greater in order to be significant). At the same time, an additional element of comparability is introduced, to take the baseline into account, as is the comparison of relative improvement, aspect in which we have also found sound results, with better results (higher effectiveness of intervention) in G1. By the other hand, it has been stated that the danger inherent in conducting small samples is that potentially effective interventions will be judged as ineffective simply because of the inability to detect statistically significant and clinically important differences [14]. However, this would affect the intervention in the G1 group, which is the one with the smallest sample.

Greater effectiveness of internal assessments

Improvement in quality means a positive change in the habitual procedures. This will be favoured if there is previously a perceived necessity for it, which habitually requires the influence of peer leaders, clinical evidence, or both things [15]. The internal QA programme and external feedback contributed to clinical evidence by providing the centres with the results of their assessments, showing a wide margin for improvement. Both methods are useful for revealing a situation which can be improved; but necessity for change does not seem to appear spontaneously. However, the internal QA programme demonstrated to be effective in quality improvement, but not so, in a general way, in the external QA programme with feedback of clinical performance. It is interesting to emphasize, however, that feedback seems to have had effect in those centres with training in QA, compared with those who had not had previous contact with these methods and philosophy. In consequence, the knowledge of the level of performance of quality care criteria is necessary, but not sufficient to improve quality.

There are other characteristics in internal QA programmes, which bestow more effectiveness on quality improvement: these can be an increase of intrinsic motivation [16] and a more favourable framework for the improvement of decision making [17]. The potential bias introduced by selection of the quality criteria by G1 centre providers is minimizing because the validity of these criteria to assess the quality of common cold management rests upon the scientific evidence. Most doctors will arrange about the relevance of performing a physical examination to a patient with respiratory signs and symptoms in order to rule out other alternatives such as a lower respiratory infection for instance [18]. In the same way, the use of systemic antihistamines, antibiotics or corticoids are not justified for the treatment of a simple common cold [19–22].

The role of external assessments and the limitations of internal QA programmes

The typical characteristics of internal QA programmes – participation of practitioners and assessment mainly focused on the process – may occasionally have limitations that have to be taken into account in the planning of strategies to implement quality improvement. Health professionals can at first appear to be reluctant to take part in activities of quality improvement because of time constraints, the anxiety provoked by the potential repercussions of these activities (criticism on the part of other colleagues) or a misleading image of most of the quality improvement programmes (disciplinary effects, bureaucratic approach etc.) [23]. Additionally, if the evaluation is done by colleagues, they may cause natural reluctance to judge the other colleagues [24]. There can also be the difficulty of a record system with insufficient data [24]. Also, quality improvement led by practitioners may focus mainly on clinical aspects that are considered more relevant for them, unduly overlooking other aspects of quality, as are satisfaction and accessibility. But the problems of purely external programmes are potentially higher. Organizations may behave as an
immune system, rejecting the introduction of new concepts of external origin. This characteristic may contribute to the poor effectiveness of external QA programmes.

External controls should facilitate the running and working of the internal programmes, through support, supervision and provision of comparative quality reports between different centres. Internal programmes seem to be more effective in quality improvement than economical incentives used in some external programmes [25]. In 1956, Lembcke had already defined external audit as that in which an agency that is not involved with the evaluated service checks periodically the precision and complete execution of internal audits [26]. Previously, the first rules for the voluntary authorization of hospitals of the American College of Surgeons in 1918 (external programme) included the requirement that the individual practitioners had to revise and analyse their clinical practice [27] at regular intervals. All these examples underline the necessity of assuring the existence of internal programmes, as suggested by the results of our study.

The cost of poor quality and the opportunity to invest in QA

In addition to the cost of the needed QA activities to deal with it, the cost of poor quality includes all kinds of wastage, unnecessary work and inadequate use of resources and its consequences [28]. The magnitude found of the cost of poor quality in our study, only in relation to the treatment prescribed for common cold patients, seems to offer a wide margin, with a high profitability, to invest in the promotion of internal programmes of QA. It is presumed that the profitability is even much higher (although more difficult to quantify) when the rest of the potential effects of inadequate treatments (adverse reactions and creation of resistance) are taken into account. It may also be much higher in other more complex pathologies.

Occasionally and for certain pathologies, quality improvement may result into an increase of the costs of care [29], by requiring additional material and human resources. This phenomenon may be wrongly viewed as a decrease of efficiency; however, from the wider perspective of quality improvement, putting resources in connection with results or products obtained, efficiency may increase because we are adding new features and better results to the service provided. In any case, the current costs of poor quality of care to numerous health problems are likely to be higher than the cost of implementing internal QA programmes. It seems that there is a very wide margin of potential profit to invest in quality.

Although our study is limited to only one health problem, our results give an empirical basis to the following conclusions:

1. The effectiveness of a quality improvement programme has more probability of success when practitioners lead the activities. The simple feedback of results of external assessments may not have any effect for the centres and health professionals who have no explicit commitment to quality.

2. Quality improvement may entail the decrease of costs of poor quality in amounts that suggest a large profitability for internal QA programmes.

3. Internal QA programmes should be incorporated into the professional culture of health institutions and be encouraged and given incentives by the system as routine activities in the health teams’ daily work, as they may have an effect on the improvement of the care provided to patients.

Acknowledgement

This study was supported by the Spanish Fund of Health Researches, Contract 94/0784, and the Department of Education and Culture Affairs of the Murcia’s Region (Grant PCT93/62).

References


