Systematic quality improvement in healthcare: clinical performance measurement and registry-based feedback
van der Veer, S.N.

Citation for published version (APA):
van der Veer, S. N. (2012). Systematic quality improvement in healthcare: clinical performance measurement and registry-based feedback

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 9

General discussion
In this thesis, we addressed the following three research topics regarding the subject of systematic quality improvement in healthcare:
1. quality improvement strategies (explored within the domain of renal replacement therapy);
2. patient experience as a clinical performance indicator (for dialysis centers);
3. the impact of registry-based feedback (examined within the context of the NICE registry).
In this final chapter, we summarize and discuss the main findings per research topic, and provide recommendations for future research.

**Quality improvement strategies in RRT care**

*Research question 1*
Which quality improvement strategies have been reported within the domain of RRT care, and what was their impact on the quality of care?

*Summary of main findings*
A large variety of quality improvement (QI) strategies and systematic QI techniques have been reported within the domain of renal replacement therapy (RRT), but due to the heterogeneity of the initiatives and the lack of rigorous evaluations no firm conclusions can be drawn with regard to their impact on RRT care.

Chapter 2 describes the results of a systematic literature review in which we identified 93 different QI initiatives that aimed to increase the uptake of best RRT practice in routine care. The majority of the initiatives combined multiple QI strategies, and used at least one systematic QI technique. Overall, patient-oriented strategies were most frequently reported. However, the number and type of strategies and techniques varied markedly between subdomains. In vascular access, almost all initiatives were multifaceted and incorporated systematic QI techniques, whereas initiatives in the domain of nutritional management mostly consisted of one element and used no QI techniques. Of the 93 initiatives, 22 were evaluated using a robust study design. Initiatives using a combination of multiple QI strategies tended to be more effective than those comprising one strategy. Initiatives using at least one systematic QI technique were not found to be more effective than those using no QI techniques at all.

*Discussing main findings – The challenge of building an evidence base for effective QI strategies*
The considerable number of publications included in the systematic reviews in Chapter 2 and Chapter 5 (on feedback by medical registries) illustrate the substantial efforts made by healthcare organizations to improve their practice. Identifying and synthesizing evidence on the effectiveness of such efforts is essential to expedite the science of successfully changing care; in the next section we discuss why this is challenging.

First of all, electronic searches of databases, such as MEDLINE, identifying evaluations of QI strategies were shown to perform suboptimally: aside from not identifying all relevant publications (lack of sensitivity), the searches yielded many irrelevant titles (lack of specificity). In both systematic reviews we applied a broad search strategy, resulting in the electronic searches identifying almost 80% of the publications we finally included; the remaining 20% were retrieved by hand searching reference lists. However, only 1-3% of all titles retrieved electronically adhered to our inclusion criteria. This low specificity may partly be due to the limited agreement on how to define and apply a common label that accurately captures QI
strategies: the terms by which they are described tend to be fuzzier than terms referring to clinical interventions. The relative lack of rigorous studies on the impact of QI strategies is a second issue. This was illustrated by the limited number of robust studies in Chapter 2, where only 25% of all evaluations were considered adequately protected against bias; in Chapter 5 half of the initiatives did not even include a baseline and follow-up measurement. Although some suggested to ease traditional evidence based medicine standards when it comes to QI publications, others advocated controlled studies as a prerequisite for separating the wheat from the chaff. Still, a robust design does not guarantee that firm conclusions on the strategy’s impact can be drawn. Most QI strategies are multifaceted, and their recipients are not single individuals but organizational units. This often makes the implementation process complicated, and implementation failures are not always obvious. This was also observed in the quantitative and qualitative evaluation of the multifaceted feedback strategy (presented in Chapters 7 and 8, respectively): the extent to which the intervention was implemented as planned varied greatly between ICUs, which formed part of our explanation for the lack of its effectiveness. Moreover, the impact and feasibility of QI strategies is influenced by context. Although variation in context is explicitly dealt with within systematic QI, it does affect generalizability of results from QI studies, especially since it remains unclear which contextual elements are most influential. In Chapter 7, for example, we suggested the fact of all ICUs being closed-format units, and the presence of the NICE data collection infrastructure as important contextual factors that contributed to the feasibility of the multifaceted feedback strategy.

Finally, in spite of a robust design, it may be difficult to make inferences on causality when evaluating the impact of QI strategies on patient outcomes. The causal link between a strategy and an outcome of care measure is often less straightforward than the link with process of care measures. This was also the case in Chapter 7, where the relationship between the feedback strategy and ICU length of stay as the primary endpoint was complex and potentially influenced by many (unknown) factors.

A third factor slowing the pace of QI science is that – in contrast to most clinical interventions – QI strategies are rarely grounded in a firm theory explaining how and why they are expected to work. This often leads to evaluations without a clear hypothesis, which in turn makes it difficult to synthesize results, and place them in the context of previous knowledge. For example, the evaluation of the feedback strategy in Chapters 7 and 8 may have yielded additional clear-cut knowledge on why performance feedback leads to change if our basic assumption would have been established in more extensive theory. Also, systematic reviews of QI strategies, such as conducted in Chapter 5, may yield more meaningful results if theory instead of clinical context is used to group strategies into conceptually coherent categories.

**Recommendations to accelerate the science of systematic quality improvement**

Based on the abovementioned challenges, we suggest the following directions for future research to accelerate the construction of a strong evidence base for effective QI strategies:

- Identification of relevant QI publications in future systematic reviews could be optimized by using validated, empirical filters as a basis for electronic searches. The recently added Medical Subject Heading (MeSH) term ‘quality improvement’ might further facilitate efficient retrieval. However, determining whether an article concerns a QI study will not be
a easy task, and accurate assignment of the new MeSH term warrants close monitoring in the coming years.

- Studies evaluating the impact of QI strategies should select a primary endpoint that reflects the strategy’s ability to change practice in addition to a (secondary) endpoint reflecting the subsequent outcome of that change. For example, a QI strategy promoting the uptake of a clinical practice guideline should be judged on its merits to increase guideline adherence rates. If patient outcomes remain unaffected despite increased adherence rates, one should first question the potential of the guideline to improve care rather than that of the QI strategy.

- Promote the use of reporting standards to ensure that QI publications contain information on the implementation process and influencing contextual factors. This will increase reproducibility and generalizability of the QI strategy to other settings. However, it is hardly possible to comply to all available reporting standards and stay within the word count limit of most peer-reviewed journals. Therefore, to facilitate concise reporting of QI studies, future research should aim to develop instruments to operationalize specific contextual factors, e.g., the Fidelity of Implementation (FOI) measure, or the Organizational Readiness to Change Assessment (ORCA) instrument. Also, there should be a focus on identifying which contextual factors are important for which QI strategies. This can be done by conducting process evaluations as an inherent part of randomized controlled QI trials (like our study described Chapter 8), synthesizing expert knowledge, or by systematically reviewing the literature.

- In addition to reporting on implementation and context, QI researchers should specify their assumption on how they expect a strategy to bring about the desired outcomes, including the theory on which it is based. As part of QI evaluations, data on intermediate measures of the implementation process, and participants’ experiences should be analyzed to check the assumption. This will contribute to further refinement of the theoretical basis for quality improvement science.

Patient experience as an indicator of the clinical performance of dialysis centers

Research question 2
How to use patient experience as an indicator of the clinical performance of dialysis centers?

Summary of main findings
The CQ index as developed in Chapter 3 is a valid and reliable instrument to measure dialysis patient experience. However, Chapter 4 shows that to correctly interpret patient experience as an indicator of dialysis center performance, one needs to adjust for several patient characteristics.

We developed two CQ index instruments (Chapter 3): one to measure patient experience with in-center hemodialysis (CHD), and one for peritoneal dialysis and home-hemodialysis (PHHD) care. The CHD instrument consisted of 42 core experience items in ten scales, of which five were reliable. For the PHHD instrument these were 31 items in nine scales; five scales were reliable. When also taking into account the priority that respondents assigned to the core experience items, the overall room for improvement appeared limited for both types of dialysis care, mainly because patients rated the experience with their dialysis care to be optimal.
Chapter 4 showed that higher ratings of dialysis centers were associated with older age, non-European ethnicity, lower educational level, no past diagnosis of malignancies, no co-morbidities, lower albumin values, and better self-rated health. Presence of a past myocardial infarction, and better self-rated health were found to be determinants of a more positive experience with the nephrologist’s care; for nurses’ care these were higher age, native Dutch ethnicity, lower educational level, lower albumin levels, and better self-rated health.

Discussing main findings – How measurable is patient experience with dialysis care?

The CQ index method used in Chapter 3 aims to capture patient experience, rather than patient satisfaction. When measuring satisfaction, patients are presumed to evaluate their treatment by comparing their personal standard with their perception of the care provided. Therefore, satisfaction is considered a subjective measure.31 The concept of patient experience is regarded more objective since it isolates actual care experience from the expectations, needs and desires patients may have before receiving treatment. This objectivity is best achieved in CQ index items referring to factual aspects of care, e.g. “did you receive information on the center's fire procedure?” Yet, many other items incorporated some degree of subjectivity (e.g., “how often did the nephrologist listen to you attentively?”), or even fully depended on the patient’s judgment (e.g., “how would you rate your center on scale from 0 to 10?”).

Besides restrictions on the attainable degree of objectivity, another limitation of instruments that measure patient experience is that pre-emptive assumptions are made about which aspects of care contribute to satisfaction.32 Within the CQ index methodology these assumptions on relevant care aspects are made by the stakeholders, and subsequently assessed by asking respondents to assign a priority to each item. However, the assigned priority does not play a formal role in selecting the core experience items to be included in the exploratory factor analysis, which forms the basis of our validation process.33 Almost half of the items originally identified as relevant aspects of care by the stakeholders were eventually excluded from this validation process either because insufficient respondents reported to have experience with those aspects, or because almost all patients reported an optimal experience. This resulted in excluding items with high priority as well as including items that may not have been rated as very important. Applying these criteria –and the fact that we did not have information on the experience of non-respondents– might have resulted in some stakeholders considering the CQ index instruments an inadequate reflection of the dialysis patient perspective. The CQ index methodology does allow for additional (high priority) aspects to be added after the validation process, but consequently these items will not pertain to any of the identified reliable scales. Hence, we can conclude from Chapter 3 that the face validity of patient experience measures is inherently jeopardized by the need for sufficient(ly variable) experience data to guarantee the reliability of such measures.

Previous research in the field of patient satisfaction showed that respondents tended to rate their care as positive, even if they did not consider their care optimal.34 35 This might be explained by people not admitting to dissatisfaction with care they chose to use, because that would suggest an inconsistency in their behavior.32 One would expect dialysis patient ratings and experience to be less susceptible to this tendency since –due to the frequent visits to the dialysis center– traveling distance will be one of the decisive factors in choosing a specific center. Still, the results from Chapters 3 and 4 show that also dialysis patients are positive about their care: only very few core experience items had more than half of the respondents reporting a suboptimal experience. Moreover, the majority of the center’s global ratings, as well as the experiences with nephrologist’s and nurses’ care, belonged to the most positive side of the
scales. Besides partly explaining the relatively small effects of patient characteristics on ratings and experience (Chapter 4), the limited variation in outcome measures also negatively affected the ability of the CQ index instruments to detect differences in patient ratings and experience between dialysis centers. A study exploring the discriminative power of the instruments concluded that—except for some scales and items—no reliable inter-center comparisons could be made based on the CQ index results.\textsuperscript{36} On the one hand, this might be explained by the instrument being insufficiently sensitive to existing differences. On the other hand, patients may consider the overall quality of delivered care to be already optimal. A larger sample size could lead to statistically significant differences, but their clinical relevance might be questionable.

**Recommendations to improve the patient perspective on dialysis care**

We have the following suggestions for future research to address the limitations regarding the objectivity, face validity, and discriminative power of the CQ index for dialysis:

- Instead of further objectifying the patient perspective, future research should focus on formalizing and incorporating the concept of patient expectation, and explore its relationship with patient characteristics and patient experience.\textsuperscript{32} This would, for example, enable investigation of the effect of improving the skills of clinicians to manage patient expectation on patient experience.\textsuperscript{37}

- To capture a more comprehensive picture of the patient perspective, quantitative methods could be complemented with qualitative methods. This would compensate for the reductionism inherent to the quantitative approach, provide more in-depth information, and facilitate the participation of groups who were missed when using a self-completion questionnaire.\textsuperscript{32} For example, the high priority aspects of dialysis care that we had to exclude due to lack of experience data can be addressed during group interviews with patients that had experience with those aspects. Future research should explore if this complementary approach yields additional opportunities for improvement.

- Although the lack of discriminative power disqualifies the CQ index as a tool within summative systems comparing Dutch dialysis centers,\textsuperscript{36} information on patient experience might still be valuable within a formative system. However, until now multifaceted feedback strategies based on patient experience data were found to be ineffective.\textsuperscript{38,39} Future research should focus on how to increase the impact of such QI strategies, e.g., by applying more sophisticated reporting tools.\textsuperscript{30} Also, more evidence is needed on which actions actually result in improved patient experience; this would provide guidance to healthcare professionals aiming to make their practice more responsive to the preferences, needs, and values of patients.

**The impact of the NICE registry feedback reports on ICU performance**

**Research question 3**

How can the impact of the NICE registry feedback reports on the quality of intensive care be increased?

**Summary of main findings**

Extending regular NICE services with more frequent and more comprehensive reports, the establishment of local QI teams, and educational outreach visits (Chapter 6) did not increase its impact on ICU patient outcomes (Chapter 7). Based on the process evaluation described in Chapter 8 we suggested that this might be partly explained by lack of meaningful benchmarks,
lack of knowledge on how to change routine practice, and insufficient allocated time and staff. Aside from this lack of effect, the feedback strategy was also shown to motivate clinicians to use performance indicators as input for quality improvement, and to form a potential first step to integrating systematic QI in daily care.

In Chapter 5 we systematically reviewed the literature and identified 50 registry feedback initiatives, covering a variety of clinical domains, and showing a large diversity in reporting formats; 22 initiatives were evaluated in a before-after study. The majority of the initiatives combined feedback reports with additional QI strategies, such as educational activities. We did not find a clear association between complementing the feedback with additional QI strategies and the impact of initiatives. Process of care measures were more often positively affected by registry-based feedback than outcome of care measures. Characteristics of the feedback itself—such as quality of the reported data, and timeliness of reporting—were most frequently suggested as factors moderating initiatives’ effectiveness.

Discussing main findings – Difficulties in translating feedback into successful QI actions

Extending the standard NICE registry reports into a multifaceted QI strategy effectively targeted part of the prospectively identified barriers to using performance feedback as a basis for systematic QI. Besides providing a stimulus and a local structure for formulating and initiating QI activities, it also increased clinicians’ trust in their own data, which they considered a prerequisite for acting on the feedback. The multifaceted strategy failed nevertheless to affect patient outcomes beyond the standard NICE registry reports. The basic underlying assumption when developing the QI strategy in Chapter 6 was that reports of inferior or inconsistent care would prompt providers to change their practice. From the results presented in Chapter 7 and Chapter 8, we learned two things with regard to this assumption: (1) healthcare providers need more sophisticated benchmarks than an unadjusted group average to determine whether their care is inconsistent; and (2) being prompted to change practice is no guarantee for knowing how to achieve this.

Previously, the national average was the only benchmark provided in the standard NICE reports. In the new feedback, ICUs could also compare their performance to the group average of similar-sized ICUs. Although it was seen as an enhancement upon the reports, QI teams often still considered it an insufficiently solid ground for deciding whether improvement was needed; this was especially true for the indicators length of ICU stay (ICU LOS) and duration of mechanical ventilation (DMV). As these measures are strongly affected by case-mix, most QI teams being confronted with a seemingly worse performance than their peers asked as a first and pertinent question: “Is our performance inconsistent because our patients are different?”. This resulted in teams spending a substantial amount of their time on investigating the characteristics of their population, but many times without being able to judge to what extent a deviant case-mix indeed accounted for the differences in performance. An additional observed effect of benchmarking based on group average was that performing close to the level of the benchmark was a reason for QI teams not to act, while in fact the average might not have corresponded to a standard of excellence.

Another limitation of the feedback reports when used as input for QI initiatives was that for a considerable part they contained information on outcome indicators, such as ICU LOS, mortality, and –to a lesser extent– DMV. In contrast to measures more closely reflecting processes of care, they were not attributable to one clinical discipline, not captured within a single care protocol, and also influenced by factors other than ICU care. This resulted in the outcome indicators often concealing the detail required to identify what went wrong, and further
drilling down was required in order to understand properly what happened, why, and what could be done to achieve improvement.\textsuperscript{41,42} This was in line with the results presented in Chapter 5. So, even when QI teams identified inconsistencies in their performance, it was often still unclear which changes were appropriate. It can be considered a weakness of the QI strategy that we did not support teams with tools for systematic problem analysis, or with suggestions for (evidence-based) actions to change practice. This line of reasoning was confirmed by the fact that the feedback on out-of-range glucose measurements—being a process of care measure—prompted most QI initiatives (Chapter 7). In all ICUs in our study glucose regulation was mainly the responsibility of nurses, was formalized in a glucose protocol, and was largely attributable to the care delivered during the ICU admission. Hence, it was often relatively clear that once inconsistent performance for this indicator had been established—actions such as educational activities aimed at nurses, and revising the glucose protocol were potentially effective in achieving improvement.

Recommendations to increase the impact of registry-based feedback on the quality of intensive care

To further increase the impact of the performance feedback provided by the NICE registry we suggest the following directions for future activities:

- To facilitate better comparison of ICUs based on the indicator feedback, the NICE registry should consider enhancing its registry reports with achievable benchmarks of care (ABCs)\textsuperscript{43} in addition to group and national averages. In a randomized controlled trial it was shown that feedback including ABCs had more impact on adherence rates to several best practices than feedback reporting only mean performance of peers.\textsuperscript{44} However, like previously discussed, adequate case-mix adjustment will be a sine qua non, especially when aiming to create meaningful achievable benchmarks for outcome indicators. This problem could be addressed by using models for case-mix adjustment. This is already the case for hospital mortality,\textsuperscript{45} but not yet for ICU LOS. Some attempts have been undertaken to adjust this outcome measure for case-mix factors.\textsuperscript{46} In turn, such attempts might facilitate, for example, prediction of a prolonged ICU stay to give clinicians a timely opportunity to explore alternatives aimed at reducing the length of the admission.\textsuperscript{47} Yet, no suitable models are currently available for the Dutch ICU population. This should be investigated in future research.

- Although outcome indicators like mortality and ICU LOS are undisputed elements of the quality of ICU care, they did not qualify as the best basis for actionable performance feedback. Out-of-range glucose measurements, incidence of severe decubitus, and the number of unplanned extubations are indicators in the current set that are much more promising in that sense. The latter two were not taken into account in our study due to data collection problems at the time of study initiation. The NICE registry should consider focusing on further improving the data quality of these indicators, and subsequently extend the feedback provided in order to facilitate formulation of appropriate actions by ICUs. Adding process indicators to the present set would be valuable, but this decision cannot be taken lightly since the road from selecting an indicator to providing feedback that is trusted by the recipients has proven to be long and winding. What might be considered is adding more structure indicators based on organizational characteristics that have an evidence-based link to improved ICU patient outcomes; the upcoming new organizational guideline from the Netherlands Society of Intensive Care (NVIC) may offer a good opportunity. The
difficulty to change such characteristics within a short time period notwithstanding, they often are actionable and easy to record.

- To support ICUs with changing their practice, the feedback reports could be enriched with suggestions for potentially effective QI actions. These suggestions can be derived from the available NVIC clinical practice guidelines, by investigating policies and practices employed by high performing ICUs, or by identifying evidence-based strategies from literature. In addition, future educational activities might include teaching the concepts and methods underlying systematic quality improvement to individual clinicians or local QI teams.

**Conclusion**

The work in this thesis shows that some aspects of care quality cannot be fully captured by one measure, that the positive impact of multifaceted registry-based feedback on clinical performance is not self-evident, and that it is difficult to extend our knowledge on how this impact can be increased. This might cause some to be disposed to conservatism about the potential of these tools to systematically improve healthcare, especially when considering that performance measurement systems add new costs to the healthcare equation. Yet, based on our work we can also conclude that it is feasible to validly measure part of a complex concept like patient experience, and that extending a registry-based feedback strategy can motivate and support clinicians to systematically improve their practice. These promising results merit further research on how to apply these tools more effectively. In addition, when looking at the myriad of performance measurement initiatives worldwide, it seems inevitable that measuring and reporting clinical performance will be increasingly institutionalized in healthcare systems, the (un)availability of evidence on their impact regardless. This further underlines the importance of rigorous investigations on how clinical performance measurement and feedback can become reliable approaches to improve the quality of healthcare.
Reference List


(44) Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA* 2001; 285:2871-2879.


