Learning from large-scale quality improvement through comparisons

JOHN ØVRETVEIT¹ AND NIEK KLAZINGA²

¹The Medical Management Centre, Karolinska Institutet, Stockholm, Sweden, and ²Department of Social Medicine, Academic Medical Centre, University of Amsterdam, P.O. Box 22660, Amsterdam NL-1100 DD, The Netherlands

Address reprint requests to: J. Øvretveit, Medical Management Centre, Karolinska Institute, Floor 5, Berzelius väg 3, Stockholm SE-171 77, Sweden. Tel: +46-31-69-39-28; Fax: +46-31-69-1777; E-mail: jovret@aol.com

Accepted for publication 5 July 2012

Abstract

Objective. To discover lessons from 10 national health and social care quality programmes in the Netherlands.

Design. A mixed-methods comparison using a ‘quantitative summarization of evidence for systematic comparison’. Each research team assessed whether there was evidence from their evaluation to support or refute 17 hypotheses about successful implementation of quality programmes. The programme managers carried out a similar assessment. Their assessments were represented as scores which made it possible to carry out a cross-case analysis to assess factors affecting the success of large-scale quality programmes.

Participants. The researchers who evaluated each of the programmes and the leaders who organized each programme.

Setting. Health and social care service organizations and national organization, which led the quality improvement programmes.

Intervention(s). This study did not make an intervention but compared experiences and evaluations of interventions carried out by national organization to health and social care service organizations to help these organizations to improve their services.

Main Outcome Measure(s). The success of the national programmes, and the learning achieved by the programme organizations and care service delivery organizations.

Results. The method provided a way to summarize and compare complex information. Common factors which appeared to influence success in implementation included understanding of political processes, leader’s influencing skills, as well as technical skills to manage projects and apply improvement and change methods.

Conclusions. Others could use a similar method to make a fast, broad level, but systematic comparison across reports of improvements or programmes. Descriptions, and then comparisons of the programmes, reveal common factors which appeared to influence success in implementation. There were groups of factors which appeared to be more important for the success of certain types of programmes. It is possible that these factors may also be important for the success of large-scale improvement programmes in other countries.

Keywords: quality improvement, quality management, leadership, quality management, Health policy, health-care system, additional specialities and/or keywords

Introduction

Many health systems are using different types of large-scale programmes to improve quality and performance as one strategy to meet the challenges of growing demands and limited resources. Examples of such programmes are accreditation schemes, programmes or reforms providing financial incentives and penalties, and schemes which publish comparative quality data [1, 2]. Some governments and independent organizations have also carried out programmes to achieve more specific improvements to services by seeking to increase the use by services of proven or promising changes to clinical practice or organization. Many of these programmes have aimed to increase the ability of services quickly to implement evidence-based treatments and improvement methods, often combining these two approaches
This paper refers to six such improvement programmes in the Netherlands [6, 7]. It describes and assesses a comparative method to discover which factors are important for success in implementing such programmes and summarizes the findings from applying this method.

Three common approaches to specific improvements are guideline implementation strategies [8], the collaborative breakthrough method [4, 9, 10] and ‘campaign’ programmes [11–13]. The last two approaches were used in six national quality improvement programmes in the Netherlands, between 2004 and 2009 and these are the subject of this paper.

The collaborative breakthrough method normally consists of a series of meetings over 1 year involving 15–100 project teams. The project teams learn about effective improvements, how to implement them locally and how to use quality methods to test and improve their version of the improvements. In effect, the collaborative approach provides a temporary infrastructure for spreading proven or promising improvements. Little is known about whether the results or activities are sustained 2 years after the end of the collaborative [14]. The campaign approach similarly involves many project teams of clinicians or care givers working to make changes, and sometimes uses the collaborative method as part of the campaign.

Systematic quality improvement methods used by these project teams in different services have been used in hospitals, and, to some extent, in primary care, but less in other health and social services. Pilot improvement projects in single services have shown some successes using these more specific improvement approaches, but little is known about the extent to which these methods can be used in a range of different ordinary services outside of hospitals and on a wide scale, their results, or how best quickly to enable many such services to use these methods [15].

Previous comparisons of large-scale specific improvement programmes over time

There are studies which have attempted to compare and draw lessons from a set of improvement programmes. Four case studies of safety programmes were carried out at four UK NHS health service sites, using retrospective ratings of safety climate and capability dimensions before and after the programme [16]. Other systematic comparisons of improvements implemented over a number of years [17–19]. Each showed the value of using a qualitative comparative case study approach to understand the implementation of improvements and to create more generalizable knowledge about conditions for success.

The Netherlands national quality programmes since 2004

The Netherlands health and social care system is a mix of public and private providers and insurers [20]. Health and social care are financed and organized into four different ‘sectors’:

- the ‘cure’ sector: hospitals and general medical practitioners in primary health care,
- the ‘care sector’: non-specialist care outside of hospitals for older and disabled people such as those provided by nursing and care homes, and by home care services,
- mental health and addiction services: specialist services, run by a variety of private providers,
- health promotion and prevention: mostly run by local authorities.

In 2004, the Netherlands started an ambitious set of programmes to improve the quality of care in these services, costing in total over €100 million by 2009. Details of each programme are reported in the evaluations of each programme, and in a meta-evaluation report [6, 7], summarized in the Supplementary data, Appendix of this paper [21]. This paper reports a comparison of 6 of 10 of the programmes, chosen for comparison because they were similar in their aim of directly improving the quality of provider services and because complete independent evaluations were available for these. The programmes were as follows:

- Quality improvement in hospitals 2004–2008 (€11 million total funding).
- Quality improvement in the care sector (plus an additional ‘National innovation’ sub programme) 2004–2011 (€30 million total funding).
- Quality improvement for dementia care 2004–2009 (€2.8 million total funding).
- Quality improvement for Mental health and addiction services 2005–2008 (€2.2 million total funding).
- Integrated diabetes care 2004–2008 (€3 million total funding), and.
- Effective illness prevention in home care services 2007–2011 (€32 million total funding).

A case study qualitative comparison was made which is not reported here, but is reported in the Supplementary data, Appendices of this paper. Further details are given in a meta-evaluation report of the programmes [7]. This paper reports the testing of hypotheses about influences which may predict success in implementation, using a method termed ‘quantitative summarization of evidence for systematic comparison’.

Method for systematic comparison of factors affecting implementation success of six programmes

The method for comparison was to examine the history of the programmes in relation to 17 hypotheses about what might predict success in implementation. This method had been used previously to test hypotheses about influences critical to success in 12 health-care innovations in Sweden [22]. To identify factors critical to the success of large-scale improvement programmes which are commonly reported in studies, a review of the literature into such programmes was made [23–25]. This was considered by the researcher teams who had carried out the evaluations of the above six programmes, before and then during a special meeting. It was revised by them to reflect their observations about factors influencing the programmes which they had evaluated, and
formulated into 17 hypotheses grouped under the three broad categories of ‘people and process’, ‘context’ and ‘content features of the innovations’ (summarized in the findings below and in 21).

A guide for assessing the evidence in relation to each hypothesis was then sent to the coordinator, programme manager and evaluation team for each of the six programmes. Each coordinator was an employee of the Netherlands health services research council, which liaised between the ministry, evaluators, and programme implementing organizations. The latter organizations were professional and service-sector associations, who were contracted to lead the implementation of the programmes in care provider services in their sectors across the country.

The guide asked these participants to use criteria to decide a score between 0 and 10 for the ‘implementation success’ of the programme. Then they were asked to take each hypothesis and consider evidence which supported or refuted the hypothesis by reflecting on their experience of the programme and by referring to documents, annual reports and the evaluations. To summarize and express their understanding of the evidence in relation to each hypothesis, they were asked to use the following score system:

+2 = strong evidence in the case confirming this hypothesis;
+1 = some evidence but not strong;
0 = no evidence for or against, but the study would have found the evidence if it was there;
−1 = some evidence refuting this hypothesis; −2 = strong evidence refuting this hypothesis, e.g. this factor played no part in the implementation progress.

The next step was to hold meetings where these scores were discussed. There was one meeting for coordinators and programme managers and another for researchers. The aim was to hear each other’s scores, and to explore the criteria each team/person was using. This allowed those present to ‘calibrate’ the meaning and specify more clearly in an agreed way which evidence was relevant, and how best to score the evidence.

The meetings considered each hypothesis, noting the scores of evidence provided by those attending for each national programme. Following this meeting, the coordinators, programme managers and researchers separately and individually reconsidered their score, submitted revised scores, and a final meeting was held to change any scores and draw conclusions from the comparisons. More details of the method are provided in the Supplementary data, Appendices to this document together with the scores provided by those who participated in the process [21].

Findings

Most common success factors

There was general agreement by participants in this process about the factors which most influenced success in implementation across the programmes. The following hypothesis all received similar and high scores, representing agreement that there was evidence from the evaluations and experience of leading the programmes that these factors most influenced programme implementation:

- Factors inside and outside the implementing organizations at provider and national levels.
- The implementation was successfully initiated and progressed by personnel at the clinical and operational levels.
- A ‘clinical champion’ (professional practitioner, e.g. nurse or doctor) who consistently spent a significant amount of time working in different ways to put the innovation into practice.
- Regular reviews and replanning of the implementation which responded to the changing situation.
- Continuity of senior management and their support at most stages to progress the improvement.
- The implementing organization had access to expertise to help with the changes.

Most of the evaluations reported that project teams in different services had difficulties in collecting data to assess their project’s progress against their change objectives. Reasons given included, lack of experience and skills in this type of data collection, lack of organizational systems for data collection and analysis, and not having the time needed to collect the data. The evaluations reported that this lack of data slowed the progress of all the programmes. However, such measurement was not rated by programme coordinators and managers as a key factor in determining implementation success.

Features of the innovations

Four features of innovations were scored as having far less influence on implementation than the factors already listed above:

- ‘A significant amount of time needed to make the change’.
- Whether the changes were ‘compatible with the organization’s existing skill mix, work practices and strategic goals’.
- ‘Lack of good evidence’, also did not appear to be a fundamental barrier to implementation improvement, especially for dementia care and for integrated diabetes care.
- ‘The innovation is/was simple to understand’ may have been a problem for the quality improvements in the care sector and integrated diabetes care, but not for others, and it did not stop implementation.

Only one feature was scored across the programmes as having some influence. This was, ‘The innovation was widely perceived to offer strong advantages compared to ‘before’ and this was important in implementation success’.

Discussion

Large-scale improvement programmes are increasingly used by health systems but there are a number of unanswered practical and scientific questions about such programmes. There are questions about their costs and effects, about
whether apparently successful programmes elsewhere can be replicated in different sectors, settings or health systems, and about how much systematic research can help answer such questions. Different approaches to studying such programmes are emerging, each with strengths and limitations for answering different questions [24–30].

All research into quality improvement interventions faces challenges in description, attribution, generalization and use [31]. By considering how these challenges were addressed in this study, the following aims to contribute to debate about how to study and provide more useful knowledge for decision-makers.

The challenges of description

Which data to collect about the intervention and context, and how to present the data are two issues facing researchers studying both small- and large-scale improvements. The SQUIRE reporting guidelines [32] give limited guidance about how much and which data to collect about large-scale programmes and their contexts. Some guidance recommends a theory-informed approach to data gathering in order to select data to gather to provide descriptions about elements of the programme and context which are thought from previous research to be relevant to the programme studied [33, 34]. Relevant generic models are provided by the PARiHS Model [27], the ORCA development of this model [35], the CFIR model [28], the MUSIQ model [29] and [30].

The six evaluations which formed the basis for this comparison each used a theory-based approach to evaluation, but some still provided limited details about the programmes and context. For those who provided more details, the information was not reported in a way which was accessible and easily usable for the comparison. It was a particular challenge for researchers to report simply and clearly the main elements of the programme and how these changed over time, in part because there were few examples to build on of how to present the data, and in part because the frameworks informing the data gathering had not been well developed, for example to specify which aspects of context could be important. Changes in the programmes also limited the validity of the assessments about each hypothesis: for example the influence of clinical champions in one programme was more important at the start than later—important knowledge about the possible relative influence of each factor at different times was obscured in the global score.

Clear, usable reporting and presentation of a programme is particularly important if decision-makers are to understand what was done in the programmes, its structure and the context so that they can assess how similar and different their situation is to the one described. Also important for description are the resource implications of the programmes—their costs and possible savings to different parties [36].

The challenges of attribution

Controlled trials are able to exclude many alternative explanations for the outcomes they study, apart from the intervention. They provide high certainty about associations between intervention and outcomes, although they do not prove causality. These designs and quasi-experimental designs can be used to answer effectiveness questions for some large-scale programmes [24–26], but many other programmes require observational, case study or programme evaluation designs for effectiveness evaluation. These designs use a variety of methods to identify and assess alternative explanations for any outcomes discovered, including triangulation of data and informed—but independent observers’ assessments [37]. All the evaluations used in the comparison reported here were challenged to show the extent to which the programme rather than other factors impacted the outcomes studied.

The challenges of generalization

If there is ‘adequate’ evidence of costs and effectiveness, then key questions for decision-makers are: is this intervention replicable in our setting? Which aspects of it can we change and still get similar results? Researchers need to describe the range of circumstances in which replication—exact or iterative—may be possible, and in which circumstances similar outcomes might be expected. Where a study is of a single quality project or programme, then a good description allows replication and can help users assess whether similar results might be expected, if the critical context factors are also described [38].

Comparative case studies of a number of quality projects within one programme can allow more valid generalization statements of whether changes can be implemented in different circumstance and whether project results could be expected in different settings. Such comparisons are more valuable if they also explore why some projects or programmes are more successful than others, rather than aggregating results from all into one global assessment. One purpose of the comparative analysis presented in this article was to assess whether hypotheses about factors important to success were specific to some programmes, or to many. The hypotheses for which there was evidence in many programmes were more likely also to apply to other large-scale national programmes, but still require testing elsewhere.

The challenges of use

Usefulness is a necessary but not sufficient condition for the use of research. Usefulness depends on the needs of the user, and whether the study is designed to meet their needs. Decision-makers at different levels need answers to questions about the resources used to make the intervention and whether any are saved, as well as whether they can implement it successfully and get similar results. For research to be of more use there is a case to be made for research designs which can answer these questions. It is possible that research designs which provide less-certain answers to a number of questions are more use to decision-makers that those which provide one answer to the one ‘does it work question?’
All the evaluations drawn on in the above study, apart from one, attempted with varying degrees of success to provide a number of answers to decision-maker’s questions. Notably, the one evaluation using a quasi-experimental design was judged by its steering committee to have provided insufficient information for users, and a process evaluation was separately commissioned which provided this information about implementation issues. To combine quasi-experimental and process evaluation designs in this way is one way to provide the range of information which decision-makers need, if the resources are available. The comparative analysis reported here aimed to provide information to decision-makers about factors which appeared to be important to success across a range of programmes and sectors.

Other comments on the method used

One purpose of the study reported here was to assess a method for testing hypotheses across cases and to assess the value of the knowledge which this method generated. The analysis performed was unique in ways which needs to be noted by others who may consider using such a method to make a fast, broad level, but systematic comparison across reports of improvements or programmes.

Each of the programmes compared had been the subject of extensive evaluations by independent research teams, who referred to their reports to examine evidence which may support or refute each hypothesis. The hypothesis testing and comparisons were made by members of the original research teams and most of the original programme implementers and coordinators also participated separately in assessments. The method could in principle be used to compare reports of programmes or to bring together those involved in such programmes, but the testing of hypothesis might have less validity without the double advantages of extensive evaluations and the experience of participants in each programme. This was because some degree of subjective judgement was needed for the global scores, especially regarding how each factors influence may have been different over 6 years.

To date most non-systematic review studies which have tried to extract lessons by comparing or reviewing different improvements or programmes have built general models of important ingredients of success [16–19]. However, these studies often combine many different programmes and interventions, and do not investigate whether certain factors are more important to certain types of interventions than others: for example whether local clinical leadership and advocacy is more important for implementing a computer physician order entry system than for implementing an ICU intervention for central line-associated blood stream infection, or for implementing a falls prevention programme in a nursing home. This criticism is less relevant in this study but still applies to an extent: the programmes were of broadly similar types in implementation approach but the content of the change were very different, and carried out in different sectors and services.

Conclusions

With the increased use of large-scale programmes to achieve rapid improvements, more knowledge is needed about their costs, effectiveness, and how successful programmes can be replicated in different settings, services and health systems. This paper reported a comparison to test 17 hypotheses about which factors were important for successful implementation of six different programmes which used similar methods in different sectors and services.

The factors consistently judged to be most influential over six programmes over 3–6 years were: access to expertise to help with the changes, clinical champions, regular reviews and replanning which responded to the changing situation, continuity of senior management and their support, and factors inside and outside of the implementing organization. These were judged more influential to successful implementation than most features of the innovation changes which were implemented, such as the compatibility, simplicity, trialability and observability of the innovation changes to be implemented in each programme. These features of the innovation changes have been reported by many previous studies to explain success of the implementation. The contrary findings of this study could be because these factors play a different role in the services and sectors in which the changes were made. It is also possible that, when viewed over a 3–6-year period, factors other than features of the innovation are more influential to successful implementation.

Both the comparative method and the evaluations on which the method drew were assessed in relation to how they address four challenges faced by many quality improvement and programme evaluation studies: challenges of description, attribution, generalization and usefulness. The limitations of the comparative method included: not being able to capture differences on the strength of influence of factors at different times in a long-term programme, possible variations in the evidence used by different participants to assess the strength of influence of a factor and of the ‘success’ of the programmes, and not including other factors in the assessment which may possibly have been influential to the success of some or all programmes. The method did however allow a broad systematic assessment of the experience and evidence about six programmes over extended time periods, and some valid lessons to be drawn which are of use in designing, implementing and studying such programmes.

Supplementary material

Supplementary material is available at INTQHC Journal online.

Acknowledgements

The authors thank members of the ZonMw community of researchers working group, which contributed to drafts of
this paper and to six-monthly discussions of the programmes. They produced the evaluations which formed one data source for the study reported in this paper. In addition, our thanks to programme managers and ZonMw coordinators who contributed to the systematic comparison and to the ZonMw research leader Dr Barbara Van de Linden for her comments on drafts of this paper and her help over the last 7 years in evaluating the overall programme.

Funding

The overview study reported in this paper was funded by ZonMw.

References


6. ZonMw ‘Documentation for the Netherlands Quality Programmes’. All programme documents, evaluations (in English) the ‘meta evaluation’ (in English) and other programme products are available in Dutch in hard copy from ZonMw. The Hague: The Netherlands Organisation for Health Research and Development (ZonMw), and from the web site www.zonmw.nl/verbeterprogrammas (31 December 2011, date last accessed).


