

Evaluation of quality improvement programmes

J Øvretveit, D Gustafson

Qual Saf Health Care 2002;11:270–275

In response to increasing concerns about quality, many countries are carrying out large scale programmes which include national quality strategies, hospital programmes, and quality accreditation, assessment and review processes. Increasing amounts of resources are being devoted to these interventions, but do they ensure or improve quality of care? There is little research evidence as to their effectiveness or the conditions for maximum effectiveness. Reasons for the lack of evaluation research include the methodological challenges of measuring outcomes and attributing causality to these complex, changing, long term social interventions to organisations or health systems, which themselves are complex and changing. However, methods are available which can be used to evaluate these programmes and which can provide decision makers with research based guidance on how to plan and implement them. This paper describes the research challenges, the methods which can be used, and gives examples and guidance for future research. It emphasises the important contribution which such research can make to improving the effectiveness of these programmes and to developing the science of quality improvement.

greater consequences for patient safety and other clinical outcomes. Yet we know little of their effectiveness or relative cost effectiveness, or how to ensure they are well implemented.

Decision makers and theorists have many questions about these programmes:

- Do they achieve their objectives and, if so, at what cost?
- Why are some more successful than others?
- What are the factors and conditions critical for success?
- What does research tell us about how to improve their effectiveness?

Some anecdotal answers come from the reports of consultants and participants, and there are theories about “critical success factors” for some types of programme. However, until recently there was little independent and systematic research about effectiveness and the conditions for effectiveness. Indeed, there was little descriptive research which documented the activities which people actually undertook when implementing a programme.

Research has made some progress in answering these questions, but perhaps not as much as was hoped, in part because of the methodological challenges. This paper first briefly notes some of the research before describing the challenges and the research designs which can be used. It finishes with suggestions for developing research in this field.

RESEARCH INTO QUALITY IMPROVEMENT PROGRAMMES

The most studied subcategory of quality programmes is hospital quality programmes, particularly US hospital total quality management programmes (TQM), later called continuous quality improvement programmes (CQI). Several non-systematic reviews have been carried out (box 1).^{2–6}

There is evidence from some studies that certain factors appear to be necessary to motivate and sustain implementation and to create conditions likely to produce results. The most commonly reported are senior management commitment, sustained attention and the right type of management roles at different levels, a focus on customer needs, physician involvement, sufficient resources, careful programme management, practical and relevant training which personnel can use immediately, and the right culture.^{4–13} These demanding conditions for success raise questions about whether the type of quality programmes which have been tried are feasible for health care. These limited conclusions appear similar across public and private, and across nations. However, there is little research for non-US clinics and hospitals, for public hospitals, or systematic comparative investigation to support this impression.

A quality programme is the planned activities carried out by an organisation or health system to improve quality. It covers a range of interventions which are more complex than a single quality team improvement project or the quality activities in one department. Quality programmes include programmes for a whole organisation (such as a hospital total quality programme), for teams from many organisations (for example, a “collaborative” programme), for external reviews of organisations in an area (for example, a quality accreditation programme), for changing practice in many organisations (for example, a practice guidelines formulation and implementation programme), and for a national or regional quality strategy which itself could include any or all of the above. These programmes create conditions which help or hinder smaller quality improvement projects.

Quality improvement programmes are new “social medical technologies” which are increasingly being applied. One study noted 11 different types of programmes in the UK NHS in a recent 3 year period.¹ They probably consume more resources than any treatment and have potentially

See end of article for authors' affiliations

Correspondence to:
Dr J Øvretveit, The Nordic
School of Public Health,
Box 12133, Goteborg,
S-40242 Sweden;
jovret@aol.com

Box 1 Non-systematic reviews of hospital quality programmes

The general conclusions of non-systematic reviews of hospital quality programmes are:

- The label given to a programme (for example, "TQM") is no guide to the activities which are actually carried out: programmes with the same name are implemented very differently at different rates, coverage, and depth in the organisation.
- Few hospitals seem to have achieved significant results and little is known about any long term results.
- Few studies describe or compare different types of hospital quality programmes, especially non-TQM/CQI programmes.
- Most studies have severe limitations (see later).

With regard to research methods, studies have tended to rely on quality specialists or senior managers for information about the programme and its impact, and to survey them once retrospectively. Future studies need to gather data from a wider range of sources and over a longer period of time. Data should also be gathered to assess the degree of implementation of the programme. Implementation should not be assumed; evidence is needed as to exactly which changes have been made and when. Outcomes need to be viewed in relation to how deeply and broadly the programme was implemented and the stage or "maturity" of the programme. To date, for most studies the lack of evidence of impact may simply reflect the fact that the programmes were not implemented, even though some respondents may say they had been. Assessing the degree of implementation could also help to formulate explanations of outcomes. There is a need for studies of organisations which are similar apart from their use of quality methods and ideas, as well the need for more studies to use the same measures—for example, of results, of culture, or of other variables. Many of these points also apply to research into other types of quality programmes.

Other quality improvement programmes

Few other types of quality improvement programmes have been systematically studied or evaluated; there are few studies of national or regional programmes such as guideline implementation or of the effectiveness of quality review or accreditation processes.¹⁴ Managers have reported that organisations which received low scores ("probation") on the US Joint Commission for Accreditation of Healthcare Organisations assessment were given high scores 3 years later but had not made substantive changes.⁶ Few studies have described or assessed the validity or value of the many comparative quality assessment systems,^{15–18} of external evaluation processes,^{19–24} or have studied national or regional quality strategies or programmes in primary health care.²⁵

More evaluation research is also being undertaken into quality improvement collaboratives. This is part of a new wave of research which is revealing more about the conditions which organisations and managers need to create in order to foster, sustain and spread effective projects and changes. Collaboratives are similar to hospital quality programmes in that they usually involve project teams, but the teams are from different organisations. The structure of the collaborative and the steps to be taken is more prescribed than most hospital quality programmes.

One study has drawn together the results of evaluations of different collaboratives.²⁶ This study provides knowledge which can be used to develop collaboratives working on other subjects, helps to understand factors critical to success, and also demonstrates other research methods which can be used to study some types of quality programmes. The study

concluded that there was some evidence that quality collaboratives can help some teams to make significant improvements quickly if the collaborative is carefully planned and managed, and if the team has the right conditions. It suggested that a team's success depended on their ability to work as a team, their ability to learn and apply quality methods, the strategic importance of their work to their home organisation, the culture of their home organisation, and the type and degree of support from management. This can help teams and their managers to decide whether they have, or can create, the conditions to be able to benefit from taking part in what can be a costly programme.

There is therefore little research into quality programmes which meets rigorous scientific criteria, but some of the research which has been done does provide guidance for decision makers which is more valid than the reports of consultants or participants. There is clearly a need for more evaluations and other types of studies of quality programmes which answer the questions of decision makers and also build theory about large scale interventions to complex health organisations or health systems. The second part of this paper considers the designs and methods which could be used in future research.

RESEARCH CHALLENGES

These interventions are difficult to evaluate using experimental methods. Many programmes are evolving, and involve a number of activities which start and finish at different times. These activities may be mutually reinforcing and have a synergistic effect if they are properly implemented: many quality programmes are a "system" of activities. Some quality programmes are implemented over a long period of time; many cannot be standardised and need to be changed to suit the situation in ways which are different from the way in which a treatment is changed to suit a patient.

The targets of the interventions are not patients but whole organisations or social groups which vary more than the physiology of an individual patient: they can be considered as complex adaptive social systems.²⁷ There are many short and long term outcomes which usually need to be studied from the perspectives of different parties. It is difficult to prove that these outcomes are due to the programme and not to something else, given the changing nature of each type of programme, their target, the environment, and the time scales involved. They are carried out over time in a changing economic, social, and political climate which influences how they are implemented.²⁸

One view is that each programme and situation is unique and no generalisations can be made to other programmes elsewhere. This may be true for some programmes, but even then a description of the programme and its context allows others to assess the relevance of the programme and the findings to their local situation. However, at present researchers do not have agreed frameworks to structure their descriptions and allow comparisons, although theories do exist about which factors are critical.

Quasi-experimental designs can be used^{29–30}; it may be possible to standardise the intervention, control its implementation, and use comparison programmes within the same environment in order to exclude other possible influences on outcomes. One issue is that many programmes are local interpretations of principles; many are not standardised specific interventions that can be replicated. Indeed, they should not be: flexible implementation for the local situation appears to be important for success.³ TQM/CQI is more a philosophy and set of principles than a specific set of steps and actions to be implemented by all organisations, although some models do come close to prescribing detailed steps.

RESEARCH DESIGNS

The difficulties in evaluating these programmes do not mean that they cannot or should not be evaluated. There are a

Box 2 A qualitative evaluation of external reviews of clinical governance

One example which illustrates the use of qualitative methods is a study of the UK government's programme of external review of clinical governance arrangements in public healthcare provider organisations.³⁵ Members of the review team as well as senior clinicians and managers were interviewed in 47 organisations before and after the review. A qualitative analysis identified themes and issues and reported common views about how the review process could be improved.

Although most interviewees thought the reviews gave a valid picture of clinical governance, much of the knowledge produced was already known to them but had not been made explicit. It concluded that major changes in policy, strategy, or direction in the organisations had not occurred as a result of the reviews, and suggested that the use of the same process for all organisations was "at best wasteful of resources and perhaps even positively harmful". This study provided the only independent description of the review process and of different stakeholders' assessments as to its value and how the process could be improved. The findings were useful to the reviewers to refine their programme. One of the limitations of the study was that it did not investigate outcomes further than the interviewees' perceptions of impact: "measuring impact reliably is difficult and different stakeholders may have quite different subjective perceptions of impact".³⁵

number of designs and methods which can and have been used: these are summarised below and discussed in detail elsewhere.^{28–34}

Descriptive case design

This design simply aims to describe the programme as implemented. There is no attempt to gather data about outcomes, but knowledgeable stakeholders' expectations of outcome and perceptions of the strengths and weaknesses of the programme can be gathered. Why is this descriptive design sometimes useful? Some quality programmes are prescribed and standardised—for example, a quality accreditation or external review. In these cases a description of the intervention activities is available which others can use to understand what was done and to replicate the intervention. However, many programmes are implemented in different ways or not described, or may only be described as principles and without a strategy. For the researcher a first description of the programme as implemented saves wasting time looking for impact further down the causal chain (for example, patient outcomes) when few or no activities have actually been implemented.

Audit design

This design takes a written statement about what people should do, such as a protocol or plan, and compares this with what they actually do. This is a quick and low cost evaluation design which is useful when there is evidence that following a programme or protocol will result in certain outcomes. It can be used to describe how far managers and health personnel follow prescriptions for quality programme interventions and why they may diverge from these prescriptions. "Audit" research of quality accreditation or review processes can help managers to develop more cost effective reviews.³⁵

Prospective before-after designs: single case or comparative

The single case prospective design gathers specific data about the target of the intervention before and after (or during) the

Box 3 Example of a theory testing comparative design

The first comprehensive studies of effectiveness of TQM/CQI programmes in health care also tried to establish which factors were critical for "success".^{8–10} The methods used in these studies were to survey 67 hospitals, some with programmes and some without, and later 61 hospitals with TQM programmes, asking questions about the programme and relating certain factors to quality performance improvement. The findings were that, after 3 years, the hospitals could not report clear evidence of results and that few had tackled clinical care processes.

A later study tested hypotheses about associations between organisation and cultural factors and performance.¹¹ Interviews and surveys were undertaken in 10 selected hospitals. Performance improvements were found in most programmes in satisfaction, market share, and economic efficiency as measured by length of stay, unit costs, and labour productivity. Interestingly, culture was only found to influence the patient satisfaction performance. It was easier for smaller hospitals with fewer complex services to implement CQI. Early physician involvement was also associated with CQI success, a finding reported in other studies.^{6,7}

This set of studies has a practical value. The findings give managers a reliable foundation for assessing whether they have the conditions which are likely to result in a successful programme. Another strength of this study was to assess the "depth" of implementation by using Baldrige or EFQM award categories.^{19,21} Limitations of the study were that: precise descriptions of the nature of the different hospital programmes were not given; only one site data gathering visit was undertaken; and less than 2 years was taken for the investigation so that the way the programmes changed and whether they were sustained could not be gauged. Follow up studies would add to our knowledge of the long term evolution of these programmes, any long term results, and explanations about why some hospitals were more successful than others.

intervention. Outcomes are considered as the differences between the before and after data collected about the target. The immediate target is the organisation and personnel; the ultimate targets are patients.

Comparative before-after designs produce stronger evidence that any outcomes were due to the programme and not to something else. If the comparable unit has no intervention, this design allows some control for competing explanations of outcomes if the units have similar characteristics and environments. These are quasi-experimental or "theory testing" designs because the researcher predicts changes to the one or more before-after variables, and then gathers the data before and after the intervention (for example, personnel attitudes towards quality) to test the prediction. However, when limited to studying only before-after (or later) differences, these designs do not generate explanations about why any changes occurred (box 2).

Retrospective or concurrent evaluation designs: single case or comparative

In these designs the researcher can use either a quasi-experimental "theory testing" approach or a "theory building" approach. An example of the former is the "prediction testing survey" design. The researcher studies previous theories or empirical research to identify theorised critical success factors—for example, sufficient resources, continuity of management, aspects of culture—and then tests these to find which are associated with successful and unsuccessful programmes (box 3).

Box 4 Example of an action evaluation comparative design

A 4 year comparative action evaluation study of six Norwegian hospitals provided evidence about results and critical factors.^{4 7 36} It gave the first detailed and long term description about what hospitals in a public system actually did and how the programmes changed over time. The study found consistencies between the six sites in the factors critical for success: management and physician involvement at all levels, good data systems, the right training, and effective project team management. A 9 year follow up is planned.

In contrast, a “theory building” approach involves the researcher in gathering data about the intervention, context, and possible effects during or after the intervention (box 4). To describe the programme as it was implemented, the researcher asks different informants to describe the activities which were actually undertaken.³⁰ The validity of these subjective perceptions can be increased by interviewing a cross section of informants, by asking informants for any evidence which they can suggest which would prove or disprove their perceptions, and by comparing data from difference sources to identify patterns in the data (box 4).^{30 32 33}

The choice of design depends on the type of quality programme (short or long term, prescribed or flexible, stable or changing), for whom the research is being undertaken, and the questions to be addressed (Was it carried out as planned? Did it achieve its objectives? What were the outcomes? What explains outcomes or success or failure?). Descriptive, audit, and single case retrospective designs are quicker to complete and are cheaper but do not give information about outcomes. Comparative outcome designs can introduce some degree of control, thus making possible inferences about critical factors if good descriptions of the programmes and their context are also provided.

IMPROVING FUTURE RESEARCH

Some of the shortcomings of research into quality programmes have been presented earlier. The five most common are:

- Implementation assessment failure: the study does not examine the extent to which the programme was actually carried out. Was the intervention implemented fully, in all areas and to the required “depth”, and for how long?
- Outcome assessment failure: the study does not assess any outcomes or a sufficiently wide range of outcomes such as short and long term impact on the organisation, on patients, and on resources consumed.
- Outcome attribution failure: the study does not establish whether the outcomes can unambiguously be attributed to the intervention, or whether something else caused the outcomes.
- Explanation failure: there is no theory or model which explains how the intervention caused the outcomes and which factors and conditions were critical.
- Measurement variability: different researchers use very different data to describe or measure the quality programme process, structure, and outcome. It is therefore difficult to use the results of one study to question or support another or to build up knowledge systematically.

Future evaluations would be improved by attention to the following:

- (1) Assessing or measuring the level of implementation of the intervention

- (2) Validating “implementation assessment”
- (3) Wider outcome assessment
- (4) Longitudinal studies
- (5) More attention to economics
- (6) Explanatory theory
- (7) Common definitions and measures
- (8) Tools to predict and explain programme effectiveness

Assessing or measuring the level of implementation of the intervention

Studies need to assess how “broadly” the programme penetrated the organisation (did it reach all parts?), how “deeply” it was applied in each part, and for how long it was applied. One of the first rules of evaluation is “assume nothing has been implemented—get evidence of what has been implemented, where and for how long”.³⁰ There is no point looking for outcomes until this has been established. Instruments for assessing “stage of implementation” or “maturation” need to be developed such as the adaptation of the Baldrige criteria used in the study by Shortell *et al*⁷ or other instruments.

Validating “implementation assessment”

Survey responses are one data source for assessing level of implementation and are useful for selecting organisations for further studies. However, these responses need to be gathered from a cross section of personnel, at different times, and supplemented by site visits and other data sources to improve validity.

Wider outcome assessment

With regard to short term impact, data need to be gathered from a wide cross section of organisational personnel and other stakeholders and from other data sources. Most studies also need to gather data about long term outcomes and to assess carefully the extent to which these outcomes can be attributed to the programme. The outcome data to be gathered should be determined by a theory predicting effects, which builds on previous research, or in terms of the specified objectives of the programme, and these links should be made clear in the report.

Longitudinal studies

Retrospective single surveys provide data which is of limited use. We need more prospective studies which follow the dynamics of the programme over long timescales. Many future studies will need to investigate both the intervention and the outcomes over an extended period of time. Very little is known about whether these programmes are continued and how they might change, or about long term outcomes.

More attention to economics

No studies have assessed the resources consumed by a quality improvement programme or the resource consequences of the outcomes. The suspected high initial costs of implementation would look different if more was known about the costs of sustaining the programme and about the possible savings and economic benefits.³⁷ Long term evaluations may also uncover more outcomes, benefits, or “side effects” which are not discovered in short studies.

Explanatory theory

For hospital programmes there is no shortage of theories about how to implement them and the conditions needed for success, but few are empirically based. For both practical and scientific reasons, future studies need to test these theories or build theories about what helps and hinders implementation at different stages, and about how the intervention produces any discovered outcomes. For other types of quality programmes there is very little theory of any type. Innovation

Box 5 Steps for studying a quality improvement programme

The methods used depend on who the research is for (the research user), the questions to be addressed, and the type of programme. An example of one action evaluation research strategy is presented here.^{30 36}

- Conceptualise the intervention. At an early stage, form a simple model of the component parts of the programme and of the activities carried out at different times. This model can be built up from programme documents or any plans or descriptions which already exist, or from previous theories about the intervention.
- Find and review previous research about similar programmes and make predictions. Identify which factors are suggested by theory or evidence to be critical for the success of the programme. Identify which variables have been studied before and how data were collected.
- Identify research questions which arise out of previous research and/or which are of interest to the users of the research.
- Consider whether the intervention can be controlled in its implementation (would people agree to follow a prescribed approach or have they done so if it is a retrospective study?). If not, design part of the study to gather data to describe the programme as implemented and to assess the level of implementation. Consider whether comparisons could be made with similar or non-intervention sites—for example, to help exclude competing explanations for outcomes or to discover assisting and hindering factors.
- Plan methods to use to investigate how the programme was actually carried out, the different activities performed, and to assess the level of implementation. Gather data about the sequence of activities and how the programme changed over time. Use documentary data sources, observation, interviews, or surveys as appropriate describing how informants or other data sources were selected and possible bias. Note differences between the planned programme and the programme in action, and participants' explanations for this as well as other explanations.
- Plan methods to gather data about the effects of the programme on providers and patients if possible. Data may be participants' subjective perceptions, or more objective before and after data (for example, complaints, clinical outcomes), or both. Use data collected by the programme participants to monitor progress and results if these data are valid. Consider how to capture data about unintended side effects—for example, better personnel recruitment and retention.
- Consider other explanations for discovered effects apart from the programme and assess their plausibility.
- To communicate the findings, create a model of the programme which shows the component parts over time, the main outcomes, and factors and conditions which appear to be critical in producing the outcomes. Specify the limitations of the study, the degree of certainty about the findings, and the answers to the research questions.

adoption³⁸ and diffusion theories are one source of ideas for building explanatory theories, for understanding level of implementation, and for understanding why some organisations are able to apply or benefit more from the intervention than others.³⁸

Common definitions and measures

Most studies to date have used their own definitions and measures of effects of quality programmes. This is now limiting our ability to compare and contrast results from different evaluation studies and to build a body of knowledge.

Tools to predict and explain programme effectiveness

Future research needs to go beyond measuring effectiveness and to give decision makers tools to predict the effects of their

programmes. Decision theory models could be used to create such tools, as could tools which effectively predict the outcomes of particular improvement projects.³⁹

In addition there is a need for overviews and theories of quality improvement programmes; we have not described the full range of interventions which fall within this category and have only given a limited discussion of a few. Future research studies need to describe the range of complex large scale quality interventions increasingly being carried out and their characteristics—for example, to describe and compare national or regional quality programmes. More consideration is needed of the similarities and differences between them, of what can be learned from considering the group as a whole, and of how theories from organisation, change management, sociology, and innovation studies can contribute to building theories about these interventions (box 5).

CONCLUSIONS

Although there is research evidence that some discrete quality team projects are effective, there is little evidence that any large scale quality programmes bring significant benefits or are worth the cost. However, neither is there strong evidence that there are no benefits or that resources are being wasted. The changing and complex features of quality programmes, their targets, and the contexts make them difficult to evaluate using conventional medical research experimental evaluation methods, but this does not mean that they cannot be evaluated or investigated in other ways. Quasi-experimental evaluation methods and other social science methods can be used. These methods may not produce the degree of certainty that is produced by a triple blind randomised controlled trial of a treatment, but they can give insights into how these processes work to produce their effects.

Conclusive evidence of effectiveness may never be possible. At this stage a more realistic and useful research strategy is to describe a programme and its context and discover factors which are critical for successful implementation as judged by different parties. In a relatively short time this will provide useful data for a more "research informed management" of these programmes.

A science is only as good as its research methods. The science of quality improvement is being developed by research into how changes to organisation and practice improve patient outcomes. However, insufficient attention has been given to methods for evaluating and understanding large scale programmes for improving quality. As these programmes are increasingly used, there is particular need for studies which do not only assess effectiveness, but also examine how best to implement them.

Authors' affiliations

J Øvretveit, Professor of Health Policy and Management, The Nordic School of Public Health and The Karolinska Institute, Sweden, and The Faculty of Medicine, Bergen University, Norway
D Gustafson, Robert Ratner Professor of Industrial Engineering & Preventive Medicine, University of Wisconsin, Madison, WI 53705, USA

REFERENCES

- 1 West E. Management matters: the link between hospital organisation and quality of patient care. *Qual Health Care* 2001;**10**:40–8.
- 2 Bigelow B, Arndt M. Total quality management: field of dreams. *Health Care Manage Rev* 1995;**20**:15–25.
- 3 Motwani J, Sower V, Brasier L. Implementing TQM in the health care sector. *Health Care Manage Rev* 1996;**21**:73–82.
- 4 Øvretveit J, Aslaksen A. *The quality journeys of six Norwegian hospitals*. Oslo: Norwegian Medical Association, 1999.
- 5 Shortell S, Bennet C, Byck G. Assessing the impact of continuous quality improvement on clinical practice: what will it take to accelerate progress. *Milbank Quarterly* 1998;**76**:593–624.
- 6 Blumenthal D, Kilo C. A report card on continuous quality improvement. *Milbank Quarterly* 1998;**76**:625–48.
- 7 Øvretveit J. The Norwegian approach to integrated quality development. *J Manage Med* 2001;**15**:125–41.

- 8 **Shortell S**, O'Brien J, Hughes H, *et al* Assessing the progress of TQM in US hospitals: findings from two studies. *Quality Leader* 1994;**6**:14–17.
- 9 **Shortell M**, O'Brien J, Carman J, *et al*. Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. *Health Serv Res* 1995;**30**:377–401.
- 10 **Carman JM**, Shortell SM, Foster, RW, *et al*. Keys for successful implementation of total quality management in hospitals. *Health Care Manage Rev* 1996;**21**:48–60.
- 11 **Boerster H**, Foster E, O'Connor, *et al*. Implementation of total quality management: conventional wisdom versus reality. *Hospital Health Serv Admin* 1996;**41**:143–59.
- 12 **Gustafson D**, Hundt A. Findings of innovation research applied to quality management principles for health care. *Health Care Manage Rev* 1995;**20**:16–24.
- 13 **Gustafson D**, Risberg L, Gering D, *et al*. *Case studies from the quality improvement support system*. ACHPR Research Report 97-0022. Washington: US Department of Health and Human Services, 1997.
- 14 **Shaw C**. External assessment of health care. *BMJ* 2001;**322**:851–4.
- 15 **Thompson R**, McElroy H, Kazandjian V. Maryland hospital quality indicator project in the UK. *Qual Health Care* 1997;**6**:49–55.
- 16 **Cleveland Health Quality Choice Program (CHQCP)**. Summary report from the Cleveland Health Quality Choice Program. *Qual Manage Health Care* 1995;**3**:78–90.
- 17 **Rosenthal G**, Harper D. Cleveland health quality choice. *Jt Comm J Qual Improve* 1994;**8**:425–42.
- 18 **Pennsylvania Health Care Cost Containment Council (PHCCC)**. *Hospital effectiveness report*. Harrisburg: Pennsylvania Health Care Cost Containment Council, 1992.
- 19 **National Institute of Standards and Technology (NIST)**. *The Malcolm Baldrige national quality award 1990 application guidelines*. Gaithersburg, MD: National Institute of Standards and Technology, 1990.
- 20 **Hertz H**, Reimann C, Bostwick M. The Malcolm Baldrige National Quality Award concept: could it help stimulate or accelerate healthcare quality improvement? *Qual Manage Health Care* 1994;**2**:63–72.
- 21 **European Foundation for Quality Management (EFQM)**. *The European Quality Award 1992*. Brussels: European Foundation for Quality Management, 1992.
- 22 **Sweeney J**, Heaton C. Interpretations and variations of ISO 9000 in acute health care. *Int J Qual Health Care* 2000;**12**:203–9.
- 23 **Shaw C**. External quality mechanisms for health care: summary of the ExPeRT project on visitatie, accreditation, EFQM and ISO assessment in European Union countries. *Int J Qual Health Care* 2000;**12**:169–75.
- 24 **Øvretveit J**. Quality assessment and comparative indicators in the Nordic countries. *Int J Health Planning Manage* 2001;**16**:229–41.
- 25 **Wensing M**, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. *Int J Qual Health Care* 1994;**6**:115–32.
- 26 **Øvretveit J**. How to run an effective improvement collaborative. *Int J Health Care Qual Assur* 2002;**15**:33–44.
- 27 **Plsek P**, Wilson T. Complexity science: complexity, leadership, and management in healthcare organisations. *BMJ* 2001;**323**:746–9.
- 28 **Øvretveit J**. Evaluating hospital quality programmes. *Evaluation* 1997;**3**:451–68.
- 29 **Cook T**, Campbell D. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally, 1979.
- 30 **Øvretveit J**. *Action evaluation of health programmes and change: a handbook for a user focused approach*. Oxford: Radcliffe Medical Press, 2002.
- 31 **Øvretveit J**. *Evaluating health interventions*. Milton Keynes: Open University Press, 1998.
- 32 **Yin R**. *Case study research: design and methods*. Beverly Hills: Sage, 1994.
- 33 **Jick T**. Mixing qualitative and quantitative methods: triangulation in action. In: Van Maanen J, ed. *Qualitative methodology*. Beverly Hills: Sage, 1983.
- 34 **Ferlie E**, Gabbay J, FitzGerald F, *et al*. Evidence-based medicine and organisational change: an overview of some recent qualitative research. In Mark A, Dopson S, eds. *Organisational behaviour in healthcare: the research agenda*. London: Macmillan, 1999.
- 35 **Walshe K**, Wallace L, Freeman T, *et al*. The external review of quality improvement in healthcare organisations: a qualitative study. *Int J Qual Health Care* 2001;**13**:367–74.
- 36 **Øvretveit J**. *Integrated quality development for public healthcare*. Oslo: Norwegian Medical Association, 1999.
- 37 **Øvretveit J**. The economics of quality: a practical approach. *Int J Health Care Qual Assur* 2000;**13**:200–7.
- 38 **Rogers E**. *Diffusion of innovation*. New York: Free Press, 1983.
- 39 **Gustafson D**, Cats-Baril W, Alemei F. *Systems to support health policy analysis*. Ann Arbor: Health Administration Press, University of Michigan, 1992.