How to Use an Article About Quality Improvement

Eddy Fan, MD
Andreas Laupacis, MD, MSc
Peter J. Pronovost, MD, PhD
Gordon H. Guyatt, MD, MSc
Dale M. Needham, MD, PhD

Quality improvement (QI) attempts to change clinician behavior and, through those changes, lead to improved patient outcomes. The methodological quality of studies evaluating the effectiveness of QI interventions is frequently low. Clinicians and others evaluating QI studies should be aware of the risk of bias, should consider whether the investigators measured appropriate outcomes, should be concerned if there has been no replication of the findings, and should consider the likelihood of success of the QI intervention in their practice setting and the costs and possibility of unintended effects of its implementation. This article complements and enhances existing Users' Guides that address the effects of interventions—Therapy, Harm, Clinical Decision Support Systems, and Summarizing the Evidence guides—with an emphasis on issues specific to QI studies. Given the potential for widespread implementation of QI interventions, there is a need for robust study methods in QI research.

JAMA. 2010;304(20):2279-2287 www.jama.com

CLINICAL SCENARIO

You, the medical director of an intensive care unit (ICU), discover that mortality has increased for patients with sepsis. You are considering a quality improvement (QI) initiative to improve the care and outcomes of your patients. However, you are concerned that many QI studies have weak designs and poor data quality and often overestimate potential benefits. Before beginning, you decide to identify and evaluate existing QI interventions.

THE SEARCH

You perform a literature search using PubMed and identify a before-after study evaluating an educational QI program for sepsis in 59 medical-surgical ICUs in Spain.[1] This program trained clinicians to recognize and treat severe sepsis based on evidence-based guidelines from the Surviving Sepsis Campaign.[2] The program implemented 2 guideline-based treatment bundles: a resuscitation bundle (6 tasks started at sepsis recognition and completed within 6 hours) and a management bundle (4 tasks completed within 24 hours).

Compared with the 2-month preintervention period, adherence to the guidelines in patients with sepsis improved for both the resuscitation bundle (5.3% vs 10.0%, $P < .001$) and the management bundle (10.9% vs 15.7%, $P = .001$) during the 4-month postimplementation period. Hospital mortality decreased (44.0% vs 39.7%, $P = .04$). One-year follow-up in a subset of participating ICUs revealed that adherence to the resuscitation bundle returned to baseline, but management bundle adherence and hospital mortality remained similar to those measures in the postintervention period. Your next step is to critically appraise the report.

CME available online at www.jamaarchivescme.com and questions on p 2306.

QUALITY IMPROVEMENT—AN OVERVIEW

Problems with the quality and safety of patient care are common.[3,4] Patients frequently do not receive evidence-based treatments,[5] and more than 9% of hospitalized patients are harmed by adverse events.[6] Quality improvement interventions attempt to change clinician behavior and, through those changes, lead to more consistent, appropriate, and efficient application of established clinical interventions, resulting in improved care and patient outcomes.[7] The likelihood of behavior change, rather than the efficacy of the interventions used, is generally at or near equipoise in QI research.

Traditional clinical research (eg, evaluating the efficacy of therapies) typically evaluates interventions delivered in well-controlled environments, increasing the likelihood that patients

Author Affiliations: Division of Pulmonary and Critical Care Medicine (Drs Fan and Needham) and Departments of Anesthesiology and Critical Care Medicine (Dr Pronovost) and Physical Medicine and Rehabilitation (Dr Needham), Johns Hopkins University, Baltimore, Maryland; Interdepartmental Division of Critical Care Medicine (Dr Fan) and Faculty of Medicine (Dr Laupacis), University of Toronto, Toronto, Ontario, Canada; Keenan Research Center, Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, Ontario (Dr Laupacis); and Departments of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario (Dr Guyatt).
Corresponding Author: Eddy Fan, MD, Division of Pulmonary and Critical Care Medicine, Johns Hopkins Medical Institutions, 1830 E Monument St, Fifth Floor, Baltimore, MD 21287 (eddy.fan@jhmi.edu).
Users' Guides to the Medical Literature Section Editor: Drummond Rennie, MD, Deputy Editor, *JAMA*.

receive the intervention.[3,4] In contrast, QI interventions are often designed to enhance the implementation of proven therapies and use data routinely collected in clinical practice. As a result, QI efforts may not always be considered research[8]; if designated as such, they may qualify for expedited review by an institutional review board that may agree to waiving informed consent because the study does not expose patients to any risks beyond those involved in standard clinical practice.[9]

Quality improvement interventions are frequently context-dependent, complex, and iterative, seeking to address barriers to and facilitators of QI.[10] When high-quality evidence has securely established substantial net benefit of existing therapies, measuring the processes of care—ie, the incremental implementation of therapies—may be sufficient to establish the benefit of QI interventions. When the net benefit of therapies is less securely established, measurement of improved patient-important outcomes is necessary to establish the benefit of QI interventions.

Quality Improvement as a Science

Randomized controlled trials (RCTs) provide an optimal strategy to reduce bias associated with prognostic differences between groups or with determinants of outcome that change over time. Publications in the QI field represent a heterogeneous literature[11] in which reports of local experience with a high risk of bias are much more frequent than reports of rigorously designed experiments.[12] Likely because of their lower risk of bias, rigorous RCTs are less likely to report significant improvements than are observational studies.[13]

Anecdotal QI reports can generate new hypotheses, fuel innovative changes, and motivate clinicians to change.[14] If no dissemination of a local QI intervention to other settings is planned, lower-quality evidence may be acceptable. However, when the apparent benefits of an intervention are widely publicized, spurious findings can result in harm, poor use of limited resources, or both. Therefore, most QI studies must be rigorously designed, conducted, and evaluated.[14-16]

Some QI initiatives have prematurely advocated the dissemination of interventions based on randomized trials with important limitations. For example, the use of β-blockers in noncardiac surgery was adopted as a target for QI efforts prior to evidence that this intervention may increase all-cause mortality and stroke.[17,18] Studies with both a low risk of bias and a high degree of applicability outside of the original research setting are needed before QI interventions are widely disseminated.[12,16]

Link to Other Users' Guides

This article complements and enhances existing Users' Guides that address the effects of interventions—Therapy, Harm, Clinical Decision Support Systems, and Summarizing the Evidence guides[19]—with an emphasis on issues specific to QI studies (Box 1). Furthermore, this guide focuses on individual studies, which, as other Users' Guides point out, should be interpreted in the context of all relevant high-quality evidence (ie, Summarizing the Evidence, Making Sense of Variability in Study Results, When to Believe a Subgroup Analysis); it also complements the Standard for Quality Improvement Reporting Excellence (SQUIRE) guidelines.[20]

Box 1. Users' Guide for an Article Assessing Quality Improvement (QI)

Are the results valid?

Did intervention and control groups start with the same prognosis?

Were patients randomized? *If not, did the investigators use an alternative design that minimizes the risk of bias?*[a]

Was randomization concealed?

Were patients in the study groups similar with respect to known prognostic factors?

If the QI intervention primarily targeted clinicians, was the clinician or the clinician group the unit of analysis?

Was data quality acceptable?

Was prognostic balance maintained as the study progressed?

To what extent was the study blinded?

Aside from the experimental intervention(s), were the groups treated equally?

Was the initial prognostic balance maintained at the completion of the study?

Was follow-up complete?

Were patients analyzed in the groups to which they were randomized *or allocated?*[a]

Was the trial stopped early?

What were the results?

How large was the treatment effect?

How precise was the estimate of the treatment effect?

How can I apply the results?

If the QI study focused on a process of care, what was the quality of evidence that the process improves patient-important outcomes?

Was follow-up sufficiently long?

Is the QI intervention exportable to my site?

Were all patient-important outcomes considered?

Are the likely benefits worth the potential hassles, harms, and costs?

[a] Points in italics represent guides specific to QI studies.

ARE THE RESULTS VALID?

Were Patients Randomized? If Not, Did the Investigators Use an Alternative Design That Minimizes the Risk of Bias?

The observational design of most QI studies may reflect events outside the control of the researchers (eg, change caused by a new policy) or the impracticability of randomization (eg, unwillingness to participate in a control group).[13] Such observational designs make it difficult to determine whether the QI intervention is responsible for observed changes, therefore generating low-quality evidence regarding the effects of the intervention on the outcome.[21]

Nonrandomized designs commonly used in QI studies include before-after studies (with and without concurrent controls), time series (interrupted or not), and stepped wedge designs (BOX 2).[22] Changes over time in patient populations or changes in practice unrelated to the QI intervention threaten the validity of uncontrolled before-after studies,[22] which often overestimate the magnitude of benefit.[23] For instance, investigators attributed an improvement in surgical outcomes in patients undergoing coronary artery bypass graft surgery in New York State to publicly reporting hospital and surgeon outcomes.[24] However, a subsequent investigation demonstrated that similar improvements occurred across the United States without any such intervention.[25]

Controlled before-after designs are infrequently used because of difficulty in identifying a suitable control group. However, even participants who appear well matched (eg, have similar demographics) at baseline may differ on important unmeasured factors (eg, adherence to the study intervention).

An interrupted time-series design may increase confidence in the causal link between the intervention and outcome (BOX 3). However, this design does not protect against the effects of other important events that may coincide with the study intervention. With this design, the periods must be explicitly defined, and statistical techniques may be required to account for autocorrelation (ie, the same type of data collected at adjacent time points are likely more similar than data collected at widely spaced points) to avoid overestimating treatment effects (Box 3).[22]

Box 2. Quantitative Study Designs for Quality Improvement (QI) Research[a]

Randomized Designs

1. Individual-patient randomized controlled trials
2. Cluster randomized trials

Nonrandomized Designs

1. Stepped Wedge Design
   Sequential rollout of QI intervention to study units (clinicians, organizations) over a number of periods, so that by the end of the study all participants have received the intervention. The order in which participants receive the intervention may be randomized (similar rigor to traditional randomized designs). Data are collected and outcomes measured at each point at which a new group of participants ("step") receives the QI intervention. Observed differences in outcomes between the control section of the wedge with those in the intervention section are assumed to be attributable to the intervention.
2. Time Series Design
   Attempt to detect whether a QI intervention has had an effect significantly greater than the underlying secular trend. Data are collected and outcomes measured at multiple points before and after the introduction of the QI intervention. Multiple points before the intervention allow the underlying trend and any secular effects to be estimated. Multiple points after the intervention allow the intervention effect to be estimated while accounting for the underlying secular trends. The intervention may be stopped (interrupted) and then reintroduced multiple times, or the intervention may be implemented just once.
3. Controlled Before-After Studies
   Identify a control population with characteristics and performance similar to those of the study population. Data are collected and outcomes measured in both the study and control populations before and after the introduction of a QI intervention to the study population. Observed differences between groups in the postintervention period or in change scores (from baseline in each group) are assumed to be attributable to the intervention.
4. Uncontrolled Before-After Studies
   Outcomes measured before and after the introduction of a QI intervention in the same study setting. Observed differences in the outcomes are assumed to be attributable to the intervention.

[a]Designs presented in order of methodological strength.

Box 3. Interrupted Time Series Design: Example

A study of thoracic surgery compared patient outcomes between a baseline period vs a postintervention period, after introduction of a clinical pathway for postoperative management. The authors compared outcomes during the baseline vs postintervention periods, reporting significant improvements.[26] However, reanalysis revealed a statistically significant preintervention trend, and time-series regression techniques demonstrated no significant differences after the intervention.[27]

Statistical process control is another common method used to analyze variations in the performance of a process over time. Variations may include improved performance in response to a QI intervention that, over time, will stabilize at a new improved level.[28]

Studies using a stepped wedge design introduce the QI intervention to participants sequentially so that, by the end of the study, all participants are exposed to the intervention (BOX 4).[30] The order in which the intervention is introduced may be randomized, further increasing evidence quality.

Some nonrandomized studies—if designed, conducted, and analyzed appropriately—may provide robust results.[16,31,32] Statistical methods (eg, regression analysis) to account for confounding variables (ie, prognostic factors that bias results because they are associated with both the QI intervention and the outcome) may strengthen observational studies.[32] When RCTs are used in QI, they are often pragmatic designs evaluating whether the QI intervention is effective among broadly defined patient groups receiving care in real-world settings (BOX 5).[22,36]

Box 4. Stepped Wedge Design: Example

Following a national UK recommendation to implement critical care outreach teams (CCOTs), one hospital undertook a stepped wedge trial evaluating the effects on hospital mortality and length of stay.[29] The CCOTs were introduced over 32 weeks, with pairing of wards to match important patient characteristics. One ward from each pair was randomized to earlier CCOT introduction, with usual care occurring in the other paired ward until subsequent CCOT introduction, allowing a matched comparison across 8 pairs of wards. The timing of CCOT initiation across ward pairs was randomly determined and phased in over time, with introduction of CCOTs in an additional ward pair at 4-month intervals. This study demonstrated that the CCOT intervention reduced in-hospital mortality vs usual care (odds ratio, 0.52 [95% confidence interval, 0.32-0.85]).

Box 5. Pragmatic Randomized Controlled Trial of a Quality Improvement Intervention: Example

In a study targeted at adult primary care patients with type 2 diabetes, investigators randomized 511 patients from 46 clinicians to receive usual care vs shared (patient-clinician) access to a Web-based, electronic diabetes tracker monitoring 13 indicators (eg, blood pressure, glycated hemoglobin [HbA_{1c}] level), as well as providing clinical advice to improve diabetes care.[33] This pragmatic randomized controlled trial was conducted among community-based clinicians to evaluate effectiveness in the setting in which most patients with diabetes receive care. The intervention group had significantly more checks of diabetic indicators than the usual-care group at 6 months (difference, 1.27 [95% confidence interval, 0.79 to 1.75]) and experienced significant improvements in blood pressure and levels of HbA_{1c}. However, HbA_{1c} level may be a poor surrogate for patient-important outcomes (ie, randomized trials of intensive therapy to achieve low HbA_{1c} targets failed to show reductions in stroke or cardiovascular death[34]); improved blood pressure is a more reliable surrogate outcome, although it too may fail to show reductions for some outcomes.[35]

Box 6. Unit of Analysis: Example

In a randomized controlled trial evaluating the effect of clinical reports encouraging use of peritoneal dialysis among patients with end-stage renal disease, 10 physicians who cared for 152 patients were randomized to the intervention or control groups.[37] The authors reported that a significantly greater number of patients started peritoneal dialysis in the group of physicians randomized to the intervention ($P = .04$). However, if the correct unit of analysis (ie, the 10 physicians, rather than the 152 patients) was used or special statistical methods were used to account for clustering of patient outcomes within physician, this result is unlikely to have reached statistical significance.[38]

If the Intervention Primarily Targeted Clinicians, Was the Clinician or Clinician Group the Unit of Analysis?

Clinicians working in the same practice, ward, or hospital share a common environment that influences practice and outcomes. Quality improvement investigators must consider this issue in their analysis. For instance, if investigators randomized hospitals to receive an intervention to improve clinician practice, a significant result may occur if data on individual clinicians' practice are analyzed without considering that individual clinicians' results are clustered (ie, physicians working in a particular hospital may practice more similarly to one another than to physicians in other hospitals).[19] Failure to appropriately consider this "unit of analysis" or "clustering" issue is common in QI studies (BOX 6).[38]

Was Data Quality Acceptable?

Although the importance of methods to control data quality is well accepted in clinical research, the same is not true in many QI studies in which data are often collected as part of routine care, without additional resources or training in research methods.[32,39] Deficiencies in data quality can invalidate study results, and the prin-

ciples of data quality should be evaluated in all phases of a QI study (BOX 7 and BOX 8).

Was Follow-up Complete?

Given the resource constraints faced in conducting most QI research, missing data are common. Missing data should be explicitly reported, because they can bias study results. If the magnitude of missing data and the potential for bias are both low in relation to the number of outcome events,[19] it may be appropriate to report the degree of missing data without explicitly addressing these data in the analysis.[32] In other situations, investigators may need to conduct sensitivity analyses to determine the potential effects of loss to follow-up. Results that do not change substantially with sensitivity analyses provide greater confidence.[32]

Back to the Original Clinical Scenario—Validity

The validity of this study is threatened by the use of an uncontrolled before-after design—an interrupted time series or stepped wedge (with randomization) study design would have less potential for bias. Quality assurance over data collection was explicitly reported, with very few missing data. The authors used regression methods to account for known imbalances between periods.

WHAT WERE THE RESULTS?

How Large and Precise Are the Effects of the Quality Improvement Intervention?

Previous Users' Guides have described common ways of expressing the effect of an intervention (eg, relative risk, risk difference) and how to evaluate its precision via a confidence interval.[19]

HOW CAN I APPLY THE RESULTS?

If the QI Study Focused on a Process of Care, What Was the Quality of Evidence That the Process Improved Patient-Important Outcomes?

Quality improvement interventions appropriately focus on process-of-care measures when processes can be accurately and feasibly measured and prior randomized trials have demonstrated that the processes improve outcomes that patients value. For instance, because RCTs have demonstrated de-

Box 7. Data Quality Control Methods for Quality Improvement (QI)

Project Design

Were the aims of the QI project clearly stated?

Were valid definitions and measurement systems reported for all important data?

Data Collection

Were staff trained, with appropriate quality assurance review, regarding data collection?

Data Management

Was there appropriate review and reporting of missing and outlier/erroneous data?

Data Analysis

Was participant flow (eg, patients, clinicians, hospitals) through the study explicitly reported (ie, number initially approached, participated, and dropped out)?

Box 8. Data Quality: Example

In a prospective, multicenter study (7688 patients) evaluating the implementation of a surgical safety checklist on patient complications,[40] data collectors at 8 international sites (including resource-poor settings) received training and supervision from local researchers on the identification, classification, and recording of process-of-care measures and complications according to the National Surgical Quality Improvement Program (NSQIP) from the American College of Surgeons. However, this training occurred only at the beginning of the quality improvement study, whereas standard NSQIP training occurs over a 1-year period, suggesting a potential limitation in training for data collectors. Furthermore, many of the complications evaluated (eg, deep-venous thrombosis) require specific diagnostic tests for accurate detection, but the proportion of patients systematically evaluated for complications as part of routine care was not reported. Thus, data quality issues may have influenced the association between the surgical safety checklist and subsequent complications.

Box 9. Quality of Evidence and Patient-Important Outcomes: Example

A prospective study evaluated the implementation and refinement of a glucose control protocol using insulin infusion (targeting blood glucose levels of 80-120 mg/dL [to convert to mmol/L, multiply by 0.0555]) for patients with hyperglycemia and sepsis in a medical intensive care unit. In 70 patients who received the protocol, 86 total hypoglycemic events were recorded, although incidence of hypoglycemia decreased from 7.6% (original protocol) to 0.3% (fourth protocol draft) as the protocol was progressively modified over the course of the study.[42] However, a subsequent randomized controlled trial enrolling critically ill patients found a significant increase in death in those randomized to target blood glucose levels of 81 to 108 mg/dL vs a target level of 180 mg/dL or lower (odds ratio, 1.14 [95% confidence interval, 1.02-1.28]).[43]

creased mortality with aspirin,[41] a QI intervention that increases the use of aspirin in patients with acute myocardial infarction leads to confidence that patients are better off, and QI researchers do not need to measure mortality.[19] In other situations, QI studies must demonstrate improvement in patient-important outcomes. Quality improvement interventions that result in more efficient or less expensive care with a neutral effect on patient-important outcomes would also be desirable. Broadly implementing interventions supported by insufficient evidence can be perilous (BOX 9).[16]

Was Follow-up Sufficiently Long?

Changes in practice may be short-lived, with many clinicians or organizations reverting to previously established routines (ie, drift in clinician behavior) once the stimulus for a new intervention is no longer present.[13] Furthermore, it often takes considerable time for groups or institutions to fully implement complex, multifaceted QI interventions. However, the median follow-up time for assessing the outcome of a QI intervention is less than 1 year in most studies,[13] which may be insufficient to determine the sustainability of the intervention. Follow-up studies of successful QI interventions should be performed to address sustainability (BOX 10). Widespread adoption of QI interventions may be unwise if the postintervention follow-up is less than 1 year.

Is the Quality Improvement Intervention Exportable to *My* Site?

The context of a QI study is key to evaluating its applicability outside of the original study site(s). Study context includes the local environment, processes, resources, leadership, culture, and traditions that cannot be accounted for in the experimental design of QI studies.[7] As a result, QI studies may involve clinicians or participants not typical of the real-world population. A clear appreciation of the local barriers to and facilitators of QI are needed before deciding if a QI study is generalizable (BOX 11).[10,47]

The context of a QI study is also important in considering the acceptability and probability of success of the intervention in different settings. Local barriers (eg, alternative therapeutic protocols) and facilitators (eg, supportive opinion leaders) may differ. For instance, in an RCT comparing usual care vs clinician education by local opinion leaders coupled with performance feedback in elderly patients following acute myocardial infarction, the QI intervention significantly increased the use of aspirin (13% vs –3%, $P = .04$) and β-blockers (31% vs 18%, $P = .02$).[47] In addition, an audit of local baseline practice may be required, because audit and feedback interventions may be most effective when baseline adherence to evidence-based practice is low.[48] Lastly, successful implementation and evaluation of a QI intervention in a number of settings (ie, replication) may help increase confidence in its generalizability. Hence, a clear understanding of context-related issues is required to understand the exportability and acceptability of the QI intervention to local practice.

Were All Patient-Important Outcomes Considered?

In reviewing the results of QI studies, one must consider whether important potential effects of the intervention were not measured.[49] Quality improvement interventions may have unintended consequences. For example, attempts to increase adherence to guidelines for colorectal cancer screening in a hospital in which many patients declined screening (because of their own preferences) resulted in decreased patient and clinician satisfaction.[50] Unintended consequences also may include effects on resource utilization and clinician behavior (TABLE 1).[51] An important unintended consequence is "crowding out" behavior, in which

Box 10. Sustainability: Examples

A cohort study examining the sustainability of a successful intervention to decrease catheter-related bloodstream infections demonstrated that infection rates remained low 36 months after the intervention.[44] In contrast, a study evaluating the implementation of a multifaceted quality improvement intervention (eg, development and dissemination of clinical guidelines supported by electronic registers, recall and reminder systems, staff education, audit, and feedback) to improve diabetes care and outcomes in Australia showed a significant improvement in service delivery (eg, clinical examinations, laboratory investigations) at 1 year (from 40% to 49%), but there was a subsequent decline in this outcome in years 2 and 3 (44% in each year).[45]

Box 11. Context: Example

Despite recommendations from professional societies and national guidelines regarding selective use of routine preoperative tests in patients with low anesthetic risk, a high number of tests were performed in a French academic hospital.[46] Previous studies demonstrated that the majority of anesthesiologists at the hospital were not complying, despite familiarity with the recommendations. Clinicians adapted the national guidelines using strategies targeted at specific organizational barriers (eg, lack of preoperative anesthesia consultation). Following this adaptation to local context, as well as active feedback regarding their practice and discussions about organizational changes, the hospital observed a sharp decrease in preoperative tests ordered for low-risk patients (80% vs 48%; $P < .05$).

gains in quality in one area occur at the expense of another (BOX 12).[51]

Are the Likely Benefits Worth the Potential Hassles, Harms, and Costs?

There are different thresholds for action based on the probabilities of benefits, harms, financial costs, and opportunity costs of a QI intervention (TABLE 2). The cost of a QI intervention (including time and effort of staff working to change clinician behavior) may be important, especially if the intervention confers only a small benefit. Moreover, it also is important to consider potential harms as well as costs resulting from not implementing a QI intervention and thus foregoing possible improvements in patient-important outcomes. Knowing such cost issues is often necessary before an organization decides to invest in that intervention, and lack of such information may delay implementation of an effective QI intervention. However, few QI studies (12%) provide economic analysis.[13]

Requiring larger, more precise (ie, narrower confidence intervals) treatment effects and rigorously con-

Table 1. Unintended Consequences of Quality Improvement (QI)

Unintended Consequence	Potential Problems	Potential Solutions
Resources	Increased costs to medical system resulting from increased direct costs of additional interventions Increased costs of data collection/information management during and after QI implementation Decreased resources allocated to other activities	Assess costs/cost-effectiveness of intervention as part of QI study Anticipate and monitor potential increased costs of the QI process; collect only essential data/information
Clinicians	Decreased attention to areas not subject to measurement in the QI intervention (ie, "crowding out" behavior) Inappropriate application of the QI intervention to ineligible patients in an effort to achieve broad success of the initiative	Consider monitoring other outcomes and clinician practices for any negative effect Ensure that the clinical area chosen for a QI intervention is the one most in need of improvement
Patients and policy makers	Access to biased or imprecise QI results impairs appropriate decision making The QI process measure may improve without change in the key patient-centered outcome QI measures or goals are inconsistent with patient preferences	Appropriate design, analysis, and reporting of QI interventions Ensure that quality measures closely match (or are good surrogate for) the desired outcome Monitor satisfaction as part of QI implementation

Box 12. Unintended Consequences: Example

A quality improvement study evaluating the effect of enhanced education and resources for medication management in depression resulted initially in improved mental health but over time resulted in reduced mental health owing to reduced coping with stress, potentially attributable to a shift away from psychological coping strategies.[52] Sustained benefits may have been realized if coping strategies were not "crowded out" by the new emphasis on medication management.

Table 2. Examples of Quality Improvement (QI) Studies and Thresholds for Decision Making in Implementing QI Interventions

	Examples of QI Studies			
Decision-Making Threshold	Multifaceted Protocol to Streamline Urgent Cardiac Catheterization and Revascularization in <90 min for Prehospital Patients With Acute ST-Segment Elevation MI	Education and Computerized Reminders for Early Administration of Empirical Antibiotics for Patients With Suspected Community-Acquired Pneumonia in the Emergency Department	Multifaceted Protocol to Improve Adherence to a Bundle of Evidence-Based Practices for Central Line Insertion to Prevent Catheter-Related Bloodstream Infections	Removing High-Concentration Intravenous Potassium Supplements From ICU Medication Box to Avoid Human Errors
Evidence for efficacy of underlying intervention in QI study	High-quality evidence of benefit[53]	High-quality evidence of benefit[54]	Moderate- to high-quality evidence of benefit[54]	No direct evidence available
Anticipated costs/harm of implementing QI intervention	High costs, potential for harm, or both	Low to moderate cost, potential for harm, or both	Low cost and potential for harm	Low cost and potential for harm
Quality of QI evidence required	Require at least moderate-quality QI evidence	Require at least moderate-quality QI evidence	Require at least low-quality QI evidence	No QI study necessary
GRADE recommendation for widespread implementation of intervention[a]	Weak (probably do it)	Strong (definitely do it)	Strong (definitely do it)	No specific recommendation

Abbreviations: GRADE, Grades of Recommendation Assessment, Development and Evaluation; ICU, intensive care unit; MI, myocardial infarction.
[a]Based on GRADE system, evaluating the balance between desirable and undesirable effects, quality of evidence, patient values and preferences, and costs (resource allocation).[19]

ducted studies may be prudent when the risks or costs of an intervention are high. When this is the case, only high confidence in an appreciable magnitude of effect can justify the intervention.[55] For QI interventions with relatively low cost and low risk, action may reasonably be taken even if a QI study has a small benefit or is supported by lower-quality evidence.[45] Large-magnitude effects are unlikely for such QI interventions. One example of a relatively low-cost and low-risk QI intervention is an observational study aimed at reducing catheter-related bloodstream infections in 103 ICUs in Michigan through use of relatively simple interventions (eg, a line insertion checklist; stocking chlorhexidine and other supplies together in a central line cart) to in turn increase use of evidence-based practices. That study demonstrated a significant decrease in the mean and median infection rate, from 7.7 and 2.7 per 1000 catheter-days at baseline to 1.4 and 0 per 1000 catheter-days in the period from 15 to 18 months after implementation ($P \leq .002$ for both comparisons).[54]

RESOLUTION OF THE CLINICAL SCENARIO

As the ICU medical director, you are faced with 2 important and related questions: did the QI intervention truly lead to improvement in the setting in which it was initially implemented, and if so, will it have a similar effect in your setting? Your research reveals that the most recent (2008) iteration of the Surviving Sepsis Campaign guidelines,[2] on which the QI intervention of this study was based, was developed using the Grades of Recommendation Assessment, Development and Evaluation (GRADE) approach.[56] Although some of the bundled interventions received a weak recommendation (eg, hydrocortisone for fluid-unresponsive septic shock), the study demonstrated a consistent increase in interventions that received a strong recommendation (eg, fluid resuscitation and vasopressors targeting a mean arterial pressure ≥65 mm Hg). Although the study shows a statistically significant decline in hospital mortality (4% absolute risk reduction) that was sustained at 1 year after implementation of the QI program, the weak study design decreases confidence in the causal association between the QI intervention and observed outcomes. Furthermore, you note that no data were reported on costs or unintended consequences. However, the substantial decrease in hospital mortality, if truly attributable to the intervention, suggests that any unintended consequences were likely small compared with the benefits.

With this Users' Guide to assist in your appraisal of this study, you decide that while confidence is low that the intervention actually improved mortality, the intervention has relatively low cost and low potential for harm. You believe the educational intervention might be successfully applied in your own hospital setting, and you will work with hospital administrators to collect data on processes of care and hospital mortality associated with this intervention in your hospital.

CONCLUSIONS

Efficacious treatments are frequently not implemented in routine clinical practice. The methodological quality of studies evaluating the effectiveness of QI interventions is frequently low. Given the potential for widespread implementation of QI interventions, there is a need for robust study methods in QI research. Clinicians and others considering implementation of QI interventions should be aware of the risk of bias in a QI study, should consider whether the investigators measured appropriate outcomes, should be concerned if there has been no replication of the findings, and should consider the likelihood of success of the QI intervention in their practice setting and the costs and possibility of unintended effects of its implementation.

Author Contributions: Dr Fan had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Fan, Laupacis, Pronovost, Guyatt, Needham.
Acquisition of data: Fan, Guyatt, Needham.
Analysis and interpretation of data: Fan, Guyatt, Needham.
Drafting of the manuscript: Fan, Pronovost, Guyatt.
Critical revision of the manuscript for important intellectual content: Fan, Laupacis, Pronovost, Guyatt, Needham.
Administrative, technical, or material support: Fan, Pronovost.
Study supervision: Fan, Guyatt, Needham.
Financial Disclosures: Dr Pronovost reported receiving grants from the Agency for Healthcare Research and Quality for reducing catheter-related bloodstream infections; receiving speaking honoraria from health systems and insurers for presentations on patient safety; and receiving an honorarium from a book on patient safety. No other authors reported disclosures.
Funding/Support: Dr Fan is supported by a Fellowship Award from the Canadian Institutes of Health Research. Dr Pronovost is supported by a K24 award from the National Institutes of Health and contracts from the Agency for Healthcare Research and Quality to improve quality of care.
Role of the Sponsors: None of the funding organizations or sponsors had any role in the design and conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.

REFERENCES

1. Ferrer R, Artigas A, Levy MM, et al; Edusepsis Study Group. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA*. 2008;299(19):2294-2303.
2. Dellinger RP, Levy MM, Carlet JM, et al; International Surviving Sepsis Campaign Guidelines Committee; American Association of Critical-Care Nurses; American College of Chest Physicians; American College of Emergency Physicians; Canadian Critical Care Society; European Society of Clinical Microbiology and Infectious Diseases; European Society of Intensive Care Medicine; European Respiratory Society; International Sepsis Forum; Japanese Association for Acute Medicine; Japanese Society of Intensive Care Medicine; Society of Critical Care Medicine; Society of Hospital Medicine; Surgical Infection Society; World Federation of Societies of Intensive and Critical Care Medicine. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med*. 2008;36(1):296-327.
3. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
4. Institute of Medicine. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academies Press; 2000.
5. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635-2645.
6. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*. 2008;17(3):216-223.
7. Batalden PB, Davidoff F. What is "quality improvement" and how can it transform healthcare? *Qual Saf Health Care*. 2007;16(1):2-3.
8. Lynn J, Baily MA, Bottrell M, et al. The ethics of using quality improvement methods in health care. *Ann Intern Med*. 2007;146(9):666-673.
9. Miller FG, Emanuel EJ. Quality-improvement research and informed consent. *N Engl J Med*. 2008;358(8):765-767.
10. Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S; SQUIRE Development Group. Publication guidelines for quality improvement in health care:

evolution of the SQUIRE project. *Qual Saf Health Care*. 2008;17(suppl 1):i3-i9.
11. Rubenstein LV, Hempel S, Farmer MM, et al. Finding order in heterogeneity: types of quality-improvement intervention publications. *Qual Saf Health Care*. 2008;17(6):403-408.
12. Thomson RG, Moss FM. QIR and SQUIRE: continuum of reporting guidelines for scholarly reports in healthcare improvement. *Qual Saf Health Care*. 2008; 17(suppl 1):i10-i12.
13. Alexander JA, Hearld LR. What can we learn from quality improvement research? a critical review of research methods. *Med Care Res Rev*. 2009;66(3): 235-271.
14. Pronovost P, Wachter R. Proposed standards for quality improvement research and publication: one step forward and two steps back. *Qual Saf Health Care*. 2006;15(3):152-153.
15. Landefeld CS, Shojania KG, Auerbach AD. Should we use large scale healthcare interventions without clear evidence that benefits outweigh costs and harms? no. *BMJ*. 2008;336(7656):1277.
16. Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med*. 2007;357(6):608-613.
17. Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet*. 2008;372(9654):1962-1976.
18. Devereaux PJ, Yang H, Yusuf S, et al; POISE Study Group. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet*. 2008; 371(9627):1839-1847.
19. Guyatt G, Rennie D, Meade M, Cook D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2nd ed.* New York, NY: McGraw-Hill Medical; 2008. http://jamaevidence.com. Accessed October 28, 2010.
20. Ogrinc G, Mooney SE, Estrada C, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care*. 2008;17(suppl 1):i13-i32.
21. Li LC, Moja L, Romero A, Sayre EC, Grimshaw JM. Nonrandomized quality improvement intervention trials might overstate the strength of causal inference of their findings. *J Clin Epidemiol*. 2009; 62(9):959-966.
22. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care*. 2003;12(1):47-52.
23. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *Am Psychol*. 1993; 48(12):1181-1209.
24. Hannan EL, Kilburn HJ Jr, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994; 271(10):761-766.
25. Ghali WA, Ash AS, Hall RE, Moskowitz MA. Statewide quality improvement initiatives and mortality after cardiac surgery. *JAMA*. 1997;277(5):379-382.
26. Zehr KJ, Dawson PB, Yang SC, Heitmiller RF. Standardized clinical care pathways for major thoracic cases reduce hospital costs. *Ann Thorac Surg*. 1998;66 (3):914-919.
27. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *Int J Technol Assess Health Care*. 2003;19(4):613-623.
28. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care*. 2007;16 (5):387-399.
29. Priestley G, Watson W, Rashidian A, et al. Introducing Critical Care Outreach: a ward-randomised trial of phased introduction in a general hospital. *Intensive Care Med*. 2004;30(7):1398-1404.
30. Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. *BMJ*. 2009; 337:a2764.
31. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? picking signal from noise. *BMJ*. 2007;334(7589):349-351.
32. Needham DM, Sinopoli DJ, Dinglas VD, et al. Improving data quality control in quality improvement projects. *Int J Qual Health Care*. 2009;21(2):145-150.
33. Holbrook A, Thabane L, Keshavjee K, et al; COMPETE II Investigators. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. *CMAJ*. 2009;181(1-2):37-44.
34. Gerstein HC, Miller ME, Byington RP, et al; Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559.
35. Chew EY, Ambrosius WT, Davis MD, et al; ACCORD Study Group; ACCORD Eye Study Group. Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*. 2010;363 (3):233-244.
36. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*. 2009;10:37.
37. Balas EA, Boren SA, Hicks LL, Chonko AM, Stephenson K. Effect of linking practice data to published evidence: a randomized controlled trial of clinical direct reports. *Med Care*. 1998;36(1):79-87.
38. Calhoun AW, Guyatt GH, Cabana MD, et al. Addressing the unit of analysis in medical care studies: a systematic review. *Med Care*. 2008;46(6):635-643.
39. Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability, and research. *Jt Comm J Qual Improv*. 1997; 23(3):135-147.
40. Haynes AB, Weiser TG, Berry WR, et al; Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360(5):491-499.
41. Baigent C, Blackwell L, Collins R, et al; Antithrombotic Trialists' (ATT) Collaboration. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. *Lancet*. 2009;373 (9678):1849-1860.
42. Clayton SB, Mazur JE, Condren S, Hermayer KL, Strange C. Evaluation of an intensive insulin protocol for septic patients in a medical intensive care unit. *Crit Care Med*. 2006;34(12):2974-2978.
43. Finfer S, Chittock DR, Su SY, et al; NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009; 360(13):1283-1297.
44. Pronovost PJ, Goeschel CA, Colantuoni E, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. *BMJ*. 2010;340:c309.
45. Bailie RS, Si D, Robinson GW, Togni SJ, D'Abbs PH. A multifaceted health-service intervention in remote Aboriginal communities: 3-year follow-up of the impact on diabetes care. *Med J Aust*. 2004;181 (4):195-200.
46. Capdenat Saint-Martin E, Michel P, Raymond JM, et al. Description of local adaptation of national guidelines and of active feedback for rationalising preoperative screening in patients at low risk from anaesthetics in a French university hospital. *Qual Health Care*. 1998;7(1):5-11.
47. Soumerai SB, McLaughlin TJ, Gurwitz JH, et al. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA*. 1998;279(17):1358-1363.
48. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? a systematic review of the effects of audit and feedback. *Qual Saf Health Care*. 2006;15(6):433-436.
49. Guyatt G, Montori VM, Devereaux PJ, Schünemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP J Club*. 2004;140 (1):A11-A12.
50. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *JAMA*. 2004;291(20):2466-2470.
51. Bardach NS, Cabana MD. The unintended consequences of quality improvement. *Curr Opin Pediatr*. 2009;21(6):777-782.
52. Wells KB, Tang L, Miranda J, Benjamin B, Duan N, Sherbourne CD. The effects of quality improvement for depression in primary care at nine years: results from a randomized, controlled group-level trial. *Health Serv Res*. 2008;43(6):1952-1974.
53. Le May MR, So DY, Dionne R, et al. A citywide protocol for primary PCI in ST-segment elevation myocardial infarction. *N Engl J Med*. 2008;358(3):231-240.
54. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med*. 2004;164 (6):637-644.
55. Berwick DM. The science of improvement. *JAMA*. 2008;299(10):1182-1184.
56. Jaeschke R, Guyatt GH, Dellinger P, et al; GRADE Working Group. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ*. 2008;337:a744.