Creating the Evidence Base for Quality Improvement Collaboratives

Brian S. Mittman, PhD

Intensive efforts are under way to improve health care quality and safety throughout the United States and abroad. Many of these efforts use the quality improvement collaborative method, an approach emphasizing collaborative learning and exchange of insights and support among a set of health care organizations. Unfortunately, the widespread acceptance and reliance on this approach are based not on solid evidence but on shared beliefs and anecdotal affirmations that may overstate the actual effectiveness of the method. More effective use of the collaborative method will require a commitment by users, researchers, and other stakeholders to rigorous, objective evaluation and the creation of a valid, useful knowledge and evidence base. Development of this evidence base will require improved conceptions of the nature of quality problems, quality improvement processes, and the types of research needed to elucidate these processes. Researchers, journal editors, and funding agencies must also cooperate to ensure that published evaluations are relevant, comprehensive, and cumulative.

Ann Intern Med. 2004;140:897-901. www.annals.org
For author affiliation, see end of text.
See related article on pp 887-896.

Deficiencies in the quality and safety of health care remain a significant concern in the United States and abroad as evidence documenting gaps between actual and recommended clinical practices continues to accumulate (1, 2). The significance of the well-recognized health care "quality chasm" has been acknowledged by a broad array of stakeholders, who have responded with efforts to identify, understand, and correct specific shortcomings in health care delivery.

The "quality improvement collaborative" (QIC) examined by Landon and colleagues in this issue (3) is arguably the health care delivery industry's most important response to quality and safety gaps; it represents substantial investments of time, effort, and funding. Largely developed and popularized by the Boston-based Institute for Healthcare Improvement (IHI) and best exemplified by IHI's Breakthrough Series collaborative program, the QIC method has been adopted on a large scale by the U.S. Health Resources and Services Administration (HRSA) (4) and the United Kingdom's National Health Service (NHS) (5). The Veterans Health Administration (6) and numerous smaller health care systems and individual hospitals and clinics worldwide have adopted the method on a smaller scale (7).

The QIC method brings together a group of participating ("collaborating") health care delivery organizations (typically between 20 and 40) and guides them in studying a specific health care quality problem, designing and implementing specific solutions, evaluating and refining these solutions, and disseminating findings to other organizations. Each participating organization is represented by a 3- or 4-person team; all teams meet together with a small faculty of experts in a series of 2 or 3 multiday collaborative "learning session" meetings, which take place over several months. During the meetings, the team members learn improvement techniques, exchange insights and advice, and generate enthusiasm and a shared sense of commitment to achieving common improvement goals and outcomes. Teams return to their organizations between learning sessions to apply their new knowledge and ideas in a Plan–Do–Study–Act framework (8). They conduct repeated cycles of quality problem diagnosis, development and implementation of small-scale improvement efforts, assessment of effects, and refinement and expansion of effective actions until desired outcomes are achieved. The QIC method is described in detail elsewhere (7, 9, 10).

Although estimates of total investment and applications of the QIC method are not available, IHI reports that "Since 1995, IHI has sponsored over 50 [Breakthrough Collaborative] projects on several dozen topics involving over 2,000 teams from 1,000 healthcare organizations" (11). The IHI tally includes a large series of collaboratives conducted under HRSA sponsorship but excludes NHS-sponsored collaboratives and numerous other collaborative efforts conducted without IHI's direct involvement.

The widespread acceptance and application of the QIC method are often questioned by observers citing the modest quantity and quality of published evidence that supports its effectiveness (12, 13). Although numerous reports of collaboratives have been published, they consist primarily of subjective or self-report assessments of QIC effects and qualitative summaries of key "lessons learned" by collaborative leaders and participants. Such first-hand reports offer important insights into the "black box" of improvement methods and processes (information generally absent from published evaluations of other quality improvement methods). However, they are incomplete without complementary, objective impact evaluations and are probably biased in favor of positive findings. This bias results from demand and supply factors. Demand-induced bias occurs because much of the published information is concentrated in management- and practitioner-oriented journals, such as *The Joint Commission Journal on Quality and Safety* and *Quality Management in Health Care,* whose

Improving Patient Care is a special section within *Annals* supported in part by the U.S. Department of Health and Human Services (HHS) Agency for Healthcare Research and Quality (AHRQ). The opinions expressed in this article are those of the authors and do not represent the position or endorsement of AHRQ or HHS.

mission and readership attract articles offering practical guidance and insights from only the successful quality improvement efforts. Supply-induced bias occurs when authors pursue quality improvement rather than research goals and document only successful collaboratives.

Methodologic weaknesses in the individual studies exacerbate distortion in the published QIC evidence base caused by publication bias. Published assessments of collaboratives generally use uncontrolled pretest–post-test designs that cannot rule out plausible alternative explanations for observed improvements, such as secular trends. (A detailed table listing published assessments of collaboratives and describing key features of the evaluation design, methods, and measures in each published article is available from the author upon request.) Although QIC authors routinely acknowledge design limitations and suggest cautious interpretation of positive findings (14), such cautions are easily overlooked. Probable errors in measurement are also pervasive. They include outcome measures that rely on participants' unvalidated self-reports or collaborative leaders' subjective ratings of readily observed phenomena (such as team enthusiasm, commitment, and adherence to the collaborative process) rather than objective measures of clinical practice or outcome change. Participants' self-reported outcome measures are typically derived from nonrigorous, highly variable measurement efforts, which often focus on short-term effects measured during or immediately following the intensive collaborative period. This results in the capture of positive effects that may be temporary and driven by Hawthorne effects rather than fundamental, lasting change. Furthermore, subjective ratings provided by collaborative participants and leaders are subject to unintentional and unrecognized biases generated by common human decision and judgment heuristics. For example, expectation biases and the phenomenon of belief perseverance combine to produce systematic overweighting of evidence and observations that confirm a priori expectations and beliefs and underweighting or discounting of evidence that does not support the effectiveness of the QIC method (15). These phenomena reinforce faith and belief in the effectiveness of the QIC method, which contributes to further accumulation of published positive findings and an escalating cycle of belief and confirmatory "evidence" (12, 16).

The apparent inconsistency between widespread belief in and use of the QIC method and the available supporting evidence heightens the importance of the study by Landon and colleagues. The authors reported on a large HRSA-sponsored collaborative that addresses the care of patients with HIV infection or AIDS. The HRSA has embraced the QIC method and has recently encouraged or mandated the participation of grantee clinics in a series of QIC projects (4). Landon and colleagues studied 44 HRSA-contracted clinics participating in the collaborative (some on a voluntary basis and some to meet contract requirements) and 25 comparison clinics. Although clinics were not randomly assigned, the authors selected comparison clinics through a careful matching process, which provided a basis for ruling out secular trends and other possible explanations of the positive effects often observed in uncontrolled pretest–post-test evaluations of collaboratives. Outcome measures were defined consistently across all sites, and comparable measures were obtained through chart review for representative patient samples from all sites. Although medical record reviewers were selected from the participating sites and the authors provide no information about chart reviewer training (or about reliability, validity checks, or other efforts to minimize measurement error), the level of rigor in the study by Landon and colleagues exceeds that of earlier evaluations of collaboratives. Any remaining bias seems more likely to produce false-positive rather than false-negative findings.

Landon and colleagues' study demonstrates small and generally insignificant pre–post improvements among both intervention and control sites in most of the 8 quality indicators measured. For the 2 indicators showing the greatest improvement (11% and 7.3%) among intervention clinics, comparable improvements were seen among control clinics as well. The small differences between the intervention and control clinics did not reach statistical significance and therefore did not preclude the attribution of observed improvement to secular trends. Several supplementary analyses compared subgroups of clinics (for example, newly funded vs. more established HRSA-funded clinics and clinics with low quality-of-care indicator scores at baseline vs. those with higher baseline scores) and subgroups of patients (for example, patients already receiving highly active antiretroviral therapy). The authors consider and rule out numerous potential explanations for their null findings, including the possibility that control clinics conducted their own improvement initiatives. They summarize many of the key design shortcomings in the existing evidence base and note that their own study would have reported (and attributed to the intervention) statistically significant improvements in 2 key measures in the absence of comparison group data. Landon and colleagues thoroughly discuss the limitations of their study but note that most seem unlikely to have produced their null results. Unfortunately, the data collected do not provide the information needed to understand and explain their findings. For example, the intervention sites may not have correctly diagnosed and understood the causes of targeted quality problems and may not have developed or fully implemented appropriate corrective actions. In addition, corrections may not have been sufficient to overcome the full spectrum of barriers to improvement. Thus, although the study by Landon and colleagues addresses many of the key methodologic problems in previously published work and offers important objective evidence on the effectiveness of the QIC method, it lacks the valuable process data reported in earlier publications (7, 14). The study's primary contribution is high-quality, objective evidence that helps

break the cycle of belief regarding the QIC method; its contributions to a more comprehensive evidence base are limited.

What are the key implications of the study by Landon and colleagues for clinical leaders and managers interested in improving quality and safety and for researchers who wish to contribute to a more useful evidence base for the QIC method and other quality improvement techniques? What steps are needed to accelerate production of valid, useful evidence and to ensure appropriate interpretation and use of this evidence in future decisions regarding use of the QIC method?

The study by Landon and colleagues and a careful reading of other available evidence on the QIC method suggest that its overall effectiveness remains highly uncertain but is probably modest. Widespread acceptance of a health care innovation in the absence of supporting evidence is not uncommon (17). The rapid diffusion of the QIC method despite the lack of supporting evidence should remind practitioners and researchers to approach quality improvement problems and methods, like other health care technologies, with a skeptical and open mind (12). Evidence to date is insufficient to reject the null hypothesis of no overall effect, that is, a normal distribution of effect sizes centered at or near zero. Without consideration of probable publication and measurement biases, published reports of quality collaboratives suggest that some organizations have achieved remarkable improvements with the QIC method while most others have attained only modest or null results (6, 10). Decisions to rely heavily on the collaborative or other methods of quality improvement should await better evidence on whether and how each method is responsible for successes. It is possible that certain highly motivated, capable organizations may achieve comparable improvements through other means.

Information is also needed to understand the reasons for QIC failures. Theory and research suggest that successful quality improvement requires a broad range of actions and supportive contextual factors (18), many of which are outside the reach of collaborative team members and their health care organization support team. The collaborative method may facilitate accurate recognition and diagnosis of quality problems. It can generate energy and commitment among team members to address these problems and provide the team with the knowledge and skills to implement solutions. However, this may not be enough to achieve substantial, lasting change. Barriers to improvement (even for relatively "simple" quality problems) remain extensive and pervasive, and required solutions are often wide-ranging and complex, requiring a spectrum of actions far beyond those that can be achieved by a QIC team. Better evidence on the operation and effects of collaboratives and reasons for variations in their effectiveness should help leaders and researchers identify situations in which the QIC method will probably be insufficient and allow them to better invest their energy and resources elsewhere. Conversely, better evidence will also allow identification of situations in which the QIC method can be a complete solution or situations in which its successful use will require support from others within or outside a health care organization.

QIC practitioners and researchers should also search thoroughly for (and try to correct) other potential weaknesses in the method and its application. These may require modest refinements in specific elements of the QIC approach, such as the number of learning sessions, or more fundamental changes, such as the selection and training of QIC leaders and participants. For example, the management theory foundations of the QIC method (9) are often overlooked in its application to health care, thereby separating the method from its supporting basic sciences of management, social psychology, sociology, economics, and political science. Effective use of the QIC method may require specialized expertise and training in these areas.

Researchers involved with QIC methods and those who use them should also recognize that the evidence goal for the QIC method is not to determine whether the method is universally "effective" or "ineffective" across diverse setting and quality problems. Research is needed to develop insights into the specific processes and mechanisms by which the QIC method and its individual components operate and to develop insights into the situational factors that facilitate or impede its acceptance, implementation, and effects. Evaluations should not assess overall QIC impacts in a black box framework but should document and elucidate improvement processes and their determinants. For example, effective evaluations will identify 1) attributes of successful faculty and faculty actions and behaviors; 2) the optimal mix of team members and site-based activities; 3) the necessary behaviors and actions of senior leaders and others within each participating organization; and 4) attributes of quality problems, organizations, and teams associated with greater success. The studies needed to provide this evidence differ significantly from the qualitative, process-oriented reports that dominate the current QIC literature and from the quantitative, impact-oriented studies (such as that by Landon and colleagues) often viewed as the methodologic gold standard in quality improvement research. Each of these approaches is insufficient to fully understand the behavioral and social phenomena central to health care quality improvement; neither approach can generate the tacit knowledge required to effectively study, refine, and apply quality improvement methods. Hybrid research approaches (13) are needed to overcome limitations of the conventional clinical research methods that are too often inappropriately applied to health care quality problems and to overcome limitations of the explicit, rule-based knowledge these methods produce.

To develop the necessary knowledge and evidence base, we require commitment and contributions from the leading proponents and users of the QIC method (for example, HRSA, IHI, and NHS), as well as researchers and

research funding agencies worldwide. Several actions are needed. First, organizations with considerable investment in the QIC method should help establish an online information clearinghouse to list and summarize completed, in-progress, and planned evaluations. This listing would facilitate awareness and application of evaluation findings. Clearinghouse contents should include published reports, supplemental details of methods and findings (often excluded from published reports because of space limitations), and evaluation instruments and tools (to facilitate replication and use in subsequent evaluations, thereby increasing comparability of evaluation results and efficiency in use of evaluation resources). Access to "public use" data sets would facilitate additional analyses and insights into the effectiveness of the QIC method. An expanded Web site that offers templates and guidance in conducting appropriate evaluations would further stimulate and support such evaluations. Valuable components of such guidance would include descriptions of quantitative and qualitative data collection and analysis methods (addressing both impact and process evaluations), suggested data sources, a compendium of suitable research designs (documenting their strengths and weaknesses), and sample institutional review board applications.

Proponents and users of the QIC method should also play a leadership role in conducting and funding comprehensive evaluations. Their involvement should be motivated by a desire to protect and leverage their investment in the QIC approach and to provide a public service (in keeping with their government or not-for-profit status) to smaller users lacking the infrastructure and resources to conduct such studies. Large QIC stakeholders, such as the HRSA, IHI, and NHS, should designate a small percentage of their investments or revenues to this purpose and should reach out to government and private foundation funding sources to supplement and leverage their contributions. Other organizations with the appropriate research orientation and infrastructure (for example, the Veterans Health Administration's Quality Enhancement Research Initiative and Kaiser Permanente's Care Management Institute) (19) should also commit to well-designed evaluations of the QIC approach.

Because individual quality improvement evaluations are typically limited in their size and generalizability, accurate interpretation and application of their findings and insights will require rigorous systematic reviews and meta-analyses. Standard analytic evidence tables must be supplemented by extensive supporting narrative that documents qualitative aspects of cross-project findings and by commentary that assesses the relevance and implications of the meta-analytic findings. Concurrent or prospective meta-analyses should be conducted to facilitate collection of consistent comparative qualitative and quantitative data. Systematic reviews and meta-analyses, like individual QIC evaluations, should endeavor to develop evidence and insights on the circumstances and factors influencing QIC effectiveness. While meeting these goals, researchers must recognize the heterogeneity of individual applications of the collaborative method and the limited value of conventional synthesis methods that are better suited to clinical evidence than evidence in the management, behavioral, and social sciences. Finally, journal editors and key researchers must help to establish standards for the design, conduct, and documentation of evaluations of quality improvement methods. Such standards will help to facilitate cumulative and comparable findings; to ensure the development and reporting of critical process as well as impact data; and to ensure access to this information for systematic reviews and meta-analyses.

The health care quality chasm remains wide and seemingly impassable. The costs of poor quality in wasted resources and in human lives and suffering remain substantial. Proponents and users of specific methods for quality improvement, particularly methods that attract the degree of trust, enthusiasm, and investment associated with the QIC method, bear an important obligation to facilitate and conduct rigorous evaluations to develop objective evidence regarding the methods' effectiveness and appropriate use.

The QIC method and others like it emphasize the importance of rigorous scientific methods, a commitment to ongoing improvement, and reliance on objective evidence and data over subjective impressions and opinion. Applying these values to the QIC method itself is no less important and should yield similar value and benefits. The result will be improved effectiveness of the method; more informed decisions on its appropriate selection and application; and, ultimately, improvements in health care quality, safety, outcomes, and efficiency.

From Veterans Affairs Quality Enhancement Research Initiative (QUERI), Washington, DC, and Veterans Affairs Greater Los Angeles Healthcare System, Sepulveda, California.

Disclaimer: The views expressed in this article are those of the author and do not necessarily represent the views of the Department of Veterans Affairs.

Grant Support: By the U.S. Department of Veterans Affairs, Health Services Research and Development Service.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Brian S. Mittman, PhD, Center for the Study of Healthcare Provider Behavior, Veterans Affairs Greater Los Angeles Healthcare System (152), 16111 Plummer Street, Sepulveda, CA 91343; e-mail, Brian.Mittman@med.va.gov.

References

1. **McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al.** The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348:2635-45. [PMID: 12826639]

2. **Institute of Medicine.** Crossing the Quality Chasm: A New Health System for

the Twenty-first Century. Washington: National Academy Pr; 2001.

3. **Landon BE, Wilson IB, McInnes K, Landrum MB, Hirschhorn L, Marsden PV, et al.** Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV study. Ann Intern Med. 2004;140: 887-96.

4. Changing lives, changing practice. Accessed at www.healthdisparities.net /BPHC_brochure.pdf on 30 March 2004.

5. National Primary Care Development Team, National Primary Care Collaborative. Accessed at www.npdt.org on 30 March 2004.

6. **Mills PD, Weeks WB.** Characteristics of successful quality improvement teams: lessons from five collaborative projects in the VHA. Jt Comm J Qual Saf. 2004;30:152-62. [PMID: 15032072]

7. **Wilson T, Berwick DM, Cleary PD.** What do collaborative improvement projects do? Experience from seven countries. Jt Comm J Qual Saf. 2003;29:85-93. [PMID: 12616923]

8. **Berwick DM.** Developing and testing changes in delivery of care. Ann Intern Med. 1998;128:651-6. [PMID: 9537939]

9. **Langley G, Nolan KM, Nolan TW, Norman CL, Provost LP.** The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco: Jossey-Bass; 1996.

10. **Ovretveit J, Bate P, Cleary P, Cretin S, Gustafson D, McInnes K, et al.** Quality collaboratives: lessons from research. Qual Saf Health Care. 2002;11: 345-51. [PMID: 12468695]

11. **Institute for Healthcare Improvement.** Accessed at www.ihi.org /newsandpublications/whitepapers/BTS.asp on 9 March 2004.

12. **Leatherman S.** Optimizing quality collaboratives. Qual Saf Health Care. 2002;11:307. [PMID: 12468688]

13. **Cretin S, Shortell SM, Keeler EB.** An evaluation of collaborative interventions to improve chronic illness care. Framework and study design. Eval Rev. 2004;28:28-51. [PMID: 14750290]

14. **Weeks WB, Mills PD, Dittus RS, Aron DC, Batalden PB.** Using an improvement model to reduce adverse drug events in VA facilities. Jt Comm J Qual Improv. 2001;27:243-54. [PMID: 11367772]

15. **Nisbett R, Ross L.** Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice Hall; 1980.

16. **Grol R.** Personal paper. Beliefs and evidence in changing clinical practice. BMJ. 1997;315:418-21. [PMID: 9277610]

17. **McKinlay JB.** From "promising report" to "standard procedure": seven stages in the career of a medical innovation. Milbank Mem Fund Q Health Soc. 1981; 59:374-411. [PMID: 6912389]

18. **Ferlie EB, Shortell SM.** Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Q. 2001;79: 281-315. [PMID: 11439467]

19. **Lomas J.** Health services research [Editorial]. BMJ. 2003;327:1301-2. [PMID: 14656811]