Chapter 1. An Introduction to the Report

Robert M. Wachter, M.D.
University of California, San Francisco

Douglas K. Owens, M.D., M.S.
VA Palo Alto Health Care System
and Stanford University

The Genesis of *Closing the Quality Gap*

Knowing is not enough; we must apply.
Willing is not enough; we must do.
— *Johann Wolfgang Von Goethe (1749-1832)*

In early 2003, the Institute of Medicine (IOM) released its report, *Priority Areas for National Action: Transforming Health Care Quality*.[1] The report listed 20 clinical topics for which best practice treatment guidelines are strongly supported by clinical evidence. Unfortunately, the report and a substantial quantity of other scientific literature show best practice implementation rates in the United States have been disappointing low, and at an annual cost of many thousands of lives.

To bring data to bear on the quality improvement opportunities laid out in the IOM's 2003 report, the Agency for Healthcare Research and Quality (AHRQ) asked the Stanford–UCSF Evidence-based Practice Center (EPC) to perform a critical analysis of the existing literature on quality improvement strategies for a number of the 20 disease and treatment priorities noted in the IOM Report. Rather than concentrating on the specific clinical practices that appear to improve health outcomes, these analyses focus on the effort of translating research into practice–identifying those activities that *increase the rate at which effective practices are applied to patient care* in real world settings. The overarching goal is one of narrowing the quality gap that is largely responsible for suboptimal health care practices and outcomes. This work also supports the recently released *National Healthcare Quality Report* (NHQR)[2] —and its companion document, the *National Healthcare Disparities Report* (NHDR).[3] Based upon earlier recommendations of the IOM,[4] Congress called upon AHRQ to deliver an annual report on the state of health care in the United States. The NHQR is intended to corroborate or refute widespread concerns related to health care quality, to document whether health care quality is stable, improving, or declining over time, and to provide national benchmarks with which individual states, health plans, and providers may compare their relative performance.

This is the first volume in a series of reports intended to support these goals. A carefully designed methodology will be applied to the scientific literature for a number of medical conditions characterized by the IOM as high-level threats to health and longevity. It is AHRQ's hope that the series will stimulate ideas for ongoing quality improvement activity nationally, as well as in individual health systems and among individual caregivers.

Origins of the Quality Movement

Although humans have long been intrigued and moved by the complex science of healing others, the science of measuring and improving the quality of delivered care is a relatively recent undertaking. Boston surgeon Ernest A. Codman (1869–1940) began his "end results system" a century ago, to track surgical outcomes and to improve surgical practice.[5] Codman's work in this area ultimately led to the creation of the Joint Commission on Accreditation of Healthcare Organizations (JCAHO).

Despite Codman's pioneering work and several other individuals and organizations whose efforts extended through the middle of the 20th Century, the science of health care quality improvement truly took root only a generation ago. Several forces catalyzed this transformation. First, medicine transcended its status as an anecdotal, non-evidence-based enterprise to one in which good data led to the discovery of improved treatment practices. For example, in the mid-1960s, 100 clinical trials were published each year. Thirty years later, that number had grown to more than 10,000.[6,7]

Second, as the public's interest and investment in the "miracles" of modern medicine grew—particularly in high technology specialties such as cardiac surgery and transplantation—so, too, did the public's demand for greater provider accountability and positive patient outcomes. Although public awareness of patient safety and quality increased with the IOM's seminal publication in 2000 of *To Err is Human: Building a Safer Health Care System,*[8] and its broader, 2001 indictment of health care quality (*Crossing the Quality Chasm*[4]) the trend had already been established. In an increasingly consumerist society, people had become less inclined to simply trust that their caregivers would deliver the highest quality care. And the public's skepticism only grew with the cost-driven growth of managed care.

Third, the expense of medical technology and the highly trained personnel needed to deliver that technology required the expansion of third-party payment systems, many of which were employer-based. These costs became a disproportionately large part of annual operating budgets and so employers, accustomed to making purchasing decisions based on *value* (quality and cost), found themselves without any information from the quality dimension of this equation. Their unwillingness to take it on faith that medicine's "product" was of the highest quality only grew with the published evidence of huge regional variations in the numbers of common procedures (coronary bypass grafting, hysterectomies, trans-urethral prostate resections) that could not be explained by differences in patient populations nor justified by differences in outcomes.[9,10] Other studies showed unacceptably high rates of "inappropriate" surgical procedures such as carotid endarterectomy,[11] further fueling the skepticism regarding the quality of health care in America and increasing demands for accountability.[12-17]

These pressures were mounting during the same time that tools for measuring the quality of evidence supporting clinical practice, such as clinical epidemiology, decision- and cost-effectiveness-analysis, meta-analysis, and the like, were becoming more robust. Driven in part by sizable congressional allocations to the National Institutes of Health and by private investments on the part of pharmaceutical companies and others, the clinical research knowledge base grew as well. The use of computer-assisted health care management systems led to the creation of large databases that could be mined to provide information on the quality of care, as large and complex clinical trials became commonplace. Before long, specialties such as cardiology, for example, were transformed. Cardiologists witnessed a shift in the cultural context, their focus

drawn away from the art of medicine and redirected toward the dozens of regularly published clinical trials and emerging evidence on best practice treatments and heart disease prevention. The probability of an American death by heart disease fell by 56 percent between 1950 and 1996. Although many factors contributed to this decline, much of the success is attributable to advances in clinical care and medical science.[18]

By the mid-1990s, the powerful influences of clinical treatment information, skepticism of the medical community's ability to ensure high-quality health care, increased consumer and purchaser knowledge, and the science of quality measurement had come together. More and more studies revealed large gaps between the findings of scientific studies and their practical implementation, even in areas of medicine where the optimal clinical approach was assured. Several large and recent studies have confirmed sizable quality of care gaps in areas spanning preventive medicine, acute and chronic care, and care of elders.[19,20-22] These and other studies emphasized the notion that research into quality health care does not, in itself, ensure the clinical patient will receive the highest quality care. A new area of inquiry—how best to translate research into practice—was born.[23-26]

Translating Research into Practice: What do we Know?

There are many reasons for the gaps that exist between the best, evidence-based understanding of high quality treatment practices, and the actual practices themselves.

First, there may be a gap in the dissemination of knowledge. The large and growing number of clinical trials underway at any given time makes it impossible for any individual physician or system of care to stay fully abreast. There is an inevitable time lag between the publication of studies that demonstrate an effective practice and its implementation. There is sometimes a need for a consensus to emerge among specialists, or a diffusion from specialists to generalists. For example, a 1994 study by Ayanian showed that cardiologists were about 10 absolute percentage points more likely to prescribe therapies known to be effective, and less likely (by roughly the same margin) to prescribe ineffective therapy, than were their generalist peers.[27] Similar lags have been demonstrated in the management of a variety of conditions,[28] ranging from peptic ulcer disease[29] to heart failure.[30]

Second, providers may be aware of a best practice, but fail to implement it because of skepticism surrounding the cost effectiveness of the practice (in terms of dollars or time needed to educate patients or adapt work processes). Or, they may have reservations regarding their treatment environment and the systems support (people or equipment) or changes in organizational culture needed to implement the practice. For example, new recommendations designed to provide improved glucose control in ICU patients[31] would mandate an increase in ICU nurse staffing to facilitate more frequent blood glucose checks. The director of critical care services at the University of California, San Francisco estimates that such increased monitoring would consume an additional 2 hours per ICU nurse shift for a typical nurse caring for two patients (Michael Gropper, MD, personal communication, 2003). Not surprisingly, many ICUs have yet to adopt this practice, despite clear and compelling evidence of the clinical benefits.

Finally, while the treatment practice may have been proven effective in a special research setting, it may not be applicable to an individual provider's setting. Clinical trials differ in many ways from real-life practice: staff members are attentive to the research protocols, personnel with specialized training may have been hired to provide additional support or patient education,

patient selection may be related to the research protocol, and additional safety measures may be built into the trial. In addition, research studies are often carried out in specialized settings (e.g., a Veterans Affairs hospital or a large academic medical center) that may bear little resemblance to the smaller treatment setting of a physician considering the practice. This gap between efficacy (how well the practice works in the research environment) and effectiveness (how well it works in clinical practice–generalized to include a wide range of treatment settings, with providers who may not be committed to or expert in its application, and a broader array of patients) has been well appreciated in recent years.[12-17]

As the quality gap has become more widely acknowledged, investigators have focused on its genesis and possible strategies for closing it. In one early analysis, Greco and Eisenberg[32] described six possible interventions to improve uptake (adoption) of improved treatment practices: education, feedback, participation by physicians in efforts to bring about change, administrative rules, financial incentives, and financial penalties.

In addition to those interventions that focus largely on the clinical behavior of individual providers (mostly physicians), more attention is being given to systematic changes in the practice environment, some of which (e.g., computerized rules and checklists, automatic stop orders) may bypass physicians entirely. A parallel movement is focusing on patients as the guardians of their own health care quality. One example cited frequently in the realm of patient safety involves patients asking their providers if they had washed their hands prior to the patient encounter.[33]

Whatever the method used to achieve the desired change, there is little doubt that the movement to base accountability and competition on metrics of quality has just begun. Business coalitions including the Pacific Business Group on Health and the Leapfrog Group are partnering with accreditation groups such as JCAHO to develop new quality-of-care standards. These standards will be made available to the public and can be used as the foundation for purchasing or payment decisions. The National Committee for Quality Assurance (NCQA) publishes its own "Report Card" for use by government agencies, employers, and consumers. Although the evidence regarding report card documents and their ability to characterize and improve overall health care quality is decidedly mixed,[34-38] public reporting and the desire to avoid negative publicity has made certain hospitals and providers eager to receive good "grades." As the case for improved quality in health care grows, so too does the realization that the best way to improve patient outcomes is a strict adherence to well-researched and respected quality improvement practices–to translate research into practice.

The Theoretic Underpinnings of Quality Improvement Efforts

Medicine has a long history of investigating what works in the clinical realm, and why. At the same time, we have a fairly limited understanding of the causal mechanisms of interventions to improve health care quality. Theories abound with regard to changing the behavior of patients, clinicians, and organizations for the better. These theories often are drawn from studies that try to isolate the effect of a single varied element or combinations of setting, interventions, and targets for change. The challenge for researchers rests in the accurate interpretation of this diverse literature regarding implementation.

In an effort to provide the reader with some context relative to the field of quality improvement (QI) implementation, this report offers a brief summary of the theoretical underpinnings that influence the development of QI interventions, as well as identifying selective

efforts that have been made to adopt and modify interventions from outside of health care. Readers interested in QI theory discussions of greater depth are encouraged to spend some time with Chapter 3, which reviews a selection of the major theories thought to influence the two dominant and parallel tracks of QI interventions: behavioral change, and the transfer or diffusion of knowledge. References to a number of pertinent theoretical models also are provided in this chapter.

An overarching theory for closing the quality gap may be neither feasible, nor desirable. Existing theories, including those from disciplines outside of health care, however, may be marshaled to design interventions for health care protocols in need of modification. Such theories have been applied in many ways—often borrowing techniques from industry such as those promoted by Juran and Deming[39-43] —with varying degrees of success. The methods generally emphasize the importance of identifying a process with less-than-ideal outcomes, measuring the key performance attributes, using careful analysis to devise a new approach, integrating the redesigned approach with the process, and reassessing performance to determine if the change in process is successful.

The mixed results produced by industry-oriented quality improvement programs (such as Total Quality Management [TQM] and Continuous Quality Improvement [CQI]) have taught managers and others the need to exercise caution before assuming that strategies drawn from other industries automatically will work in health care settings, and demonstrated that additional attention that must be given to the forces that promote desired behavioral changes among front-line workers.[44-46] These forces are an outgrowth of human needs and desires: the altruism of most health care professionals, their desire for success and peer respect, their preference for avoiding embarrassment, and the goal of financial independence, to name but a few. These inspirations have prompted a more recent movement, in which the traditional quality improvement sensibilities of programs such as TQM or CQI are coupled with more modern approaches to behavior modification, such as performance auditing and feedback. An audit often will measure provider adherence to a specific process or treatment practice, and the providers being studied will receive comparative data after the fact about their performance and how they stack up against their peers. In other types of audits, providers might receive financial rewards for their strict adherence to desired behaviors, or information regarding their performance and standing might be forwarded on to their patients (who can influence non-conforming providers to make the appropriate behavioral change, or choose to seek care elsewhere).

Remarkably, considering the enormous stakes, there has been little information written about the most effective ways to translate research into practice. Even for common disorders like diabetes, hypertension, and cancer care—areas in which research has successfully demonstrated that some best practices can save tens of thousands of lives—there has been only modest systematic study of the techniques and strategies shown to close the quality gap. Moreover, in those few areas that have benefited from such studies, little consideration has been given to crosscutting practices (i.e., how a practice that closes the quality gap in asthma, might be applicable to congestive heart failure).

What Conclusions can be Drawn from the Report's Evidence?

This report is intended to help readers assess whether the available evidence suggests that a quality improvement strategy would work in their specific practice setting, or, within their specific patient population. Three important questions should be considered:

1. Are the studies of the strategy *valid*? A study has validity (sometimes called *internal validity*) if its findings are likely to be true in the population on which the study was based. The primary determinant of validity is the design and conduct of the study.
2. For the quality improvement strategies that have been evaluated in multiple studies with sufficient validity, does the evidence indicate that the strategy is *effective*?
3. Are the conclusions of a body of evidence applicable to a practice setting or population of interest?

Careful attention has been paid to the design of each included study (Chapter 2), as a means of assisting readers to better judge study validity. Except where noted, the review has been restricted to studies that are likely to have strong validity, i.e., randomized controlled trials, well controlled before–after studies, and interrupted-time-series studies. This has been done to acknowledge an important tension in the field of quality improvement. Given the challenges and constraints of studying change in complex organizations, some authorities consider some of the most relevant QI work to be that performed "in the trenches," by front-line workers taking advantage of available resources to answer important, practical questions using simple designs (e.g., uncontrolled before-and-after studies). This point of view has relevance. However, in a report of this type, the authors placed a priority on finding and analyzing those studies with research methodologies most likely to give scientifically correct answers: randomized controlled studies, controlled before-and-after studies, and interrupted-time-series studies. They did so with the recognition that the relatively strict criteria may have led to the exclusion of some studies with potentially relevant findings.

When specific QI strategies have been evaluated in the course of multiple studies, deciding if the weight of the evidence favors the strategy can be a complex decision. To help readers make this assessment, the authors have used tables to indicate the range of results for different strategies. In those instances where studies were sufficiently similar in their design and sample size to justify combining the results, the authors used quantitative methods of analysis to synthesize their findings. When it was judged imprudent to combine studies quantitatively, the researchers made every attempt to highlight important findings and, when possible, they noted whether the findings are consistent across studies. The methods used in the course of these analyses are described in greater detail in Chapter 2.

Perhaps the most difficult challenge facing the authors of this report and its readers concerns the applicability of a study's results to a particular treatment setting or a patient population other than that used in the study itself. Studies vary in terms of the disease process considered, the population sample, the type of quality improvement intervention scrutinized, the behavior addressed by the intervention, and the time frame of the study. Each of these factors affects the applicability (sometimes called "generalizability") of the study. For example, if a study showed that audit and feedback improved prescribing practices for hypertension in a managed care treatment setting, would these findings hold true in a fee-for-service practice? Would they hold

true for diabetes care? If audit and feedback was effective in a general medicine clinic, would the same improvement strategy prove equally effective for a specialty clinic?

Caution is warranted with respect to any study's results and their applicability across settings or diseases, as the specific conditions of any user's practice are certain to differ from those of the study population. The factors with the greatest effect on the applicability of study findings are not yet known, but the final evidence report of the series will describe the EPC's findings and experience in the hope that the reader will be able to evaluate any common findings across different disease processes.

The Organizational Framework of this Series

Volume 1 contains this introduction to the series, the evidence-based methodology that unifies and underlies each of the treatment condition reports in the series (Chapter 2), and the theories thought to influence QI and implementation (Chapter 3).

Volumes 2 and 3 will review the evidence regarding the effectiveness of QI implementation practices in the treatment of diabetes and hypertension, respectively. These volumes, and those to follow, will feature the same detailed organizational framework:

Introduction – The authors identify the general background and clinical context for the disease or condition, they illustrate the primary quality gap(s) for the topic, and provide a means of benchmarking outcomes for these problems. The best treatment practices also are provided, as are the strategies for quality improvement.

Methodology – The scope of material reviewed for the topic is delimited, noting studies that have been excluded, and specifying the primary outcomes of interest. Some information pertaining to the methodologic process and analysis appears in the Methods section of Volume 1 as well as in Volume 2 (Diabetes) and Volume 3 (Hypertension). This redundancy was planned for the reader's convenience, since each of the volumes dealing with priority conditions may be read as a stand-alone analysis.

Findings Overview – A summary of the reviewed literature is provided, along with two separate analyses: one delineated by outcome and one by quality improvement strategy. An Appendix for each volume provides tables of included studies and results.

Discussion – An analysis will be included for each of the studied priority conditions, with a list of the strategies best supported by the available evidence, as well as obvious gaps and suggestions for future research.

Subsequent volumes in the series, to be produced over the next two years, will consider the evidence behind QI practices for a select number of conditions from the IOM's 2003 quality report. Evidence for the impact of individual QI practices in specific diseases or care settings will be considered in condition- or setting-specific volumes. Global analysis of the QI practices across diseases or settings will likely be addressed in the final volume in the series. The last volume also may be used to describe broad themes that emerge from the project. Finally, attempts will be made to quantify and prioritize the benefits of the various QI strategies, to the extent that the published evidence permits.

This page intentionally left blank