UNIVERSITY OF CALIFORNIA

Los Angeles

Improving Quality Improvement in Surgery:
The Role of Quality Improvement Collaboratives

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Health Policy and Management

by

Aaron Jay Dawes

2016

ABSTRACT OF THE DISSERTATION

Improving Quality Improvement in Surgery:

The Role of Quality Improvement Collaboratives

by

Aaron Jay Dawes

Doctor of Philosophy in Health Policy and Management

University of California, Los Angeles, 2016

Professor Robert H. Brook, Chair

Few topics in American medicine have generated as much interest or debate as the quality of the healthcare we receive. The United States continues to rank well below its developed peers in many international comparisons due, in part, to the persistence of medical errors, preventable deaths, and sub-optimal patient outcomes. In an effort to close these and other gaps in quality, physicians, health services researchers, and improvement experts have searched--and continue to search--for tools that can consistently and sustainably improve clinical performance. This dissertation explores one potential improvement tool: the quality improvement collaborative (QIC). QICs, which require healthcare organizations to work together on a single quality problem, were initially designed simply to disseminate evidence-based best practices, but have evolved into a more general approach for helping organizations make targeted changes to their care delivery systems. From a theoretical perspective, the QIC approach offers several

advantages over single-institution improvement strategies, including the ability to rely upon an external support system and to leverage the practical knowledge of organizations that have already made similar changes. However, the empirical data on QICs' effectiveness are mixed: a few studies suggest that hospitals achieve more significant improvements in QICs than on their own, but several randomized trials show no difference. To better understand the reasons for this inconsistency, I examined the mechanisms behind QIC function and developed a framework for understanding variability in QIC success. I then went on to explore two ways in which QICs might influence quality improvement more generally: by collecting new data or developing new risk adjustment models and by rapidly adapting an intervention from another setting to the local environment. As a whole, my research suggests that, while QICs offer a series of unique benefits to both hospitals and to the field, they are, by no means, a panacea. While there is still no magic bullet, I believe that continuing to develop the QIC approach, identifying where and when it is most effective, and integrating it into the larger armamentarium of improvement tools offers us the best chance to improve the quality of American healthcare for good.

The dissertation of Aaron Jay Dawes is approved.

Henry M. Cryer

Clifford Y. Ko

Brian S. Mittman

Jack Needleman

Robert H. Brook, Committee Chair

University of California, Los Angeles

2016

For C.P. Kelsey (1924-2015) and Herbert Kramer (1931-2015),

two men of quality in every way.

Table of Contents

Table of Tables

Table of Figures

Table of Common Abbreviations

BTS, Breakthrough Series (a model for QICs developed by the Institute for Healthcare Improvement)
CCM, Chronic Care Model (a healthcare delivery model that focuses on multidisciplinary and collaborative care for patients with chronic medical conditions)
CQI, continuous quality improvement
EMS, Emergency Medical Services Agency (a division of the Los Angeles County Department of Health Services)
ICU, intensive care unit
IHI, Institute for Healthcare Improvement
LA, Los Angeles
LACTC, Los Angeles County Trauma Consortium (a QIC)
NNECVDSG, Northern New England Cardiovascular Disease Study Group (an early QIC)
QIC, quality improvement collaborative
RCT, randomized controlled trial
SSI, surgical site infection
TBI, traumatic brain injury
VON, Vermont Oxford Network (an early QIC)

Acknowledgements

"Piglet noticed that even though he had a Very Small Heart, it could hold a rather large amount of Gratitude." -Alan Alexander Milne, *Winnie the Pooh*

I am deeply indebted to my advisors, colleagues, friends, and family, all of whom have contributed--both directly and indirectly--to this dissertation. I would like to start by thanking my chair, Bob Brook, who has tirelessly read version after version of this dissertation and, above all, has inspired it through his own efforts to improve quality over the last 40 years. Bob, thank you for always believing in me and for challenging me to make a difference. You once believed that you would witness the end to all of our gaps in quality; I can only hope to help to one day fulfill your dream.

I would also like to thank my committee members who have shared their knowledge, time, and expertise with me over these last three years. Dr. Cryer, thank you for introducing me to this project and for going well out of your way to support me at every stage. Dr. Ko, thank you for mentoring me from the beginning of residency and for teaching me the technical (and the political) side of quality measurement. Dr. Mittman, thank you for helping me explore the inner workings of QICs and for opening my eyes to the field of implementation science. Dr. Needleman, thank you for teaching me how to not only be a health services researcher, but a good one; I promise that my days of "rushing to the regression" are long behind me.

Several individuals outside of my committee also played key roles in improving this dissertation. I would like to thank Gery Ryan for his assistance with my qualitative methodology in Chapter 4 and Ronald Hays and Scott Comulada for their valuable input on my variable selection techniques in Chapter 5. I would also like to thank all of the members of the Los Angeles County Trauma Consortium. Helping to establish and develop our Consortium has been a true pleasure and one that I will never forget.

Chapter 6 is a version of Dawes AJ, Sacks GD, Cryer HG, Gruen JP, Preston C, Gorospe D, Cohen M, McArthur DL, Russell MM, Maggard-Gibbons M, and Ko CY for the Los Angeles County Trauma Consortium, "Compliance With Evidence-Based Guidelines and Interhospital Variation in Mortality for Patients With Severe Traumatic Brain Injury." *JAMA Surg*. 2015;150(10):965-972, although the chapter contains several additional analyses that were not included in the original published version. I would also like to acknowledge generous funding from the Veterans Affairs/Robert Wood Johnson Clinical Scholars Program, the Veterans Affairs Office of Academic Affiliations, the Dr. Ursula Mandel Scholarship, and the UCLA Graduate Research Mentorship Award.

Finally, I would like to thank my family and friends for supporting me through what have been some of the most stressful, but most enjoyable years in recent memory. To my parents, Paul and Betsy, and my parents (in law), Rich and Phyllis: I could not have had a better set of cheerleaders. Thank you for reminding me that there is always an end in sight. I saved my last and largest thank you for my wife, Danielle. Thank you for not only putting up with me, but supporting me through this entire process. In our darkest depths and our brightness lights, you are there to share my life. It would be nothing without you.

Aaron J. Dawes
Culver City, California
May 2016

Biographical Sketch

Aaron Jay Dawes is a physician, doctoral candidate, and health services researcher interested in improving the quality of surgical care. Dr. Dawes grew up in the San Francisco Bay Area before attending college at Princeton University in Princeton, New Jersey. While at Princeton, Dr. Dawes fulfilled both his general and pre-medical requirements, but quickly became interested in health policy. He applied and was accepted into the Woodrow Wilson School for Public and International Affairs as a sophomore.

While at the Wilson School, Dr. Dawes pursued a range of policy projects, including early childhood education in New Jersey and workforce development in South Africa. He spent the second semester of his junior year at the University of Cape Town where he participated in a working group on labor policy that culminated in a presentation to the country's finance minister. Upon returning to the United States, Dr. Dawes spent a summer serving as an intern to the health policy division of the Minority Office of the Senate Health, Education, Labor, and Pensions (HELP) Committee under Senator Edward Kennedy. While there, he worked on a variety of projects, including policies to promote the expansion of electronic health records and to improve the quality of care provided by the Indian Health Service.

As a senior at Princeton, Dr. Dawes completed an undergraduate thesis, entitled "The 'D' Stands for Disaster: A Behavioral Psychology Approach to the Medicare Part D Dilemma," under the direction of Dr. Janet Schwartz (Psychology/Public Affairs). Although it argued in favor of the new Medicare Part D plan from a traditional economic perspective, the thesis also brought to light several important design flaws from a behavioral economics perspective in an attempt to explain significant gaps in enrollment.

After graduating from Princeton, Dr. Dawes returned to the Bay Area to work as a research assistant for the Division of Pediatric Surgery at Stanford University, Stanford, California before attending medical school at Vanderbilt University School of Medicine in Nashville, Tennessee. Between his first and second years of medical school, Dr. Dawes served as an Ensign in the U.S. Public Health Service and was assigned to the Indian Health Service at Fort Defiance Indian Hospital, Fort Defiance, Arizona. While there, Dr. Dawes assisted in patient care as a part of the general pediatrics service and carried out a series of administrative duties, including updating public health readiness protocols for the entire reservation.

Upon returning to Vanderbilt, Dr. Dawes worked as a research assistant in the labs of Dr. Eric Grogan (Thoracic Surgery) and Dr. Nipun Merchant (Surgical Oncology). In addition to his research work, Dr. Dawes was an active participant in student life, serving as a tutor, student affiliate advisor, and eventually student body president in his final year.

In 2011, Dr. Dawes returned to California to begin his residency in general surgery at UCLA. After completing two years of training, he was accepted to the Veterans Affairs/Robert Wood Johnson Foundation Clinical Scholars Program and began a series of projects on surgical health services research. His primary project during the Scholars program was co-founding the Los Angeles County Trauma Consortium, a quality improvement collaborative that includes representatives from all 14 Los Angeles County trauma centers, the County's Emergency Medical Services Agency, and two local universities. After two years in the program, Dr. Dawes was awarded an additional year of funding though the Office of Academic Affiliations, Department of Veterans Affairs to continue his research on quality of care.

Upon the completion of his training, Dr. Dawes will return to his clinical residency at UCLA, with plans to pursue an additional clinical fellowship before starting his career in academic surgery. He currently lives with his wife, Danielle, and their dog, Chance, in Culver City, CA.

Chapter 1 An Introduction to Quality Improvement Collaboratives

Quality of care continues to rank among the most discussed and most researched topics in modern medicine. From only 10,000 citations in 1968 when PubMed launched its subject heading on "Quality of Care," the database has expanded to encompass over 5 million peer-reviewed articles on quality, of which over 1 million were published in the last four years alone.[1] Researchers, professional organizations, and advisory bodies, such as the Institute of Medicine, continue to publish evidence-based recommendations and clinical guidelines on an almost continual basis, updating previous versions with new findings, sources, and methods. Beyond its significance to research, quality has entered the day-to-day lives of clinicians and healthcare organizations across the country. U.S. physician practices currently spend more than $15.4 billion per year collecting and reporting data for external quality measures.[2] Hospitals, hospital systems, and even physicians themselves are becoming used to having their performance compared to that of others in their region. *U.S. News and World Report*, famous for its ranking of colleges and universities, publishes yearly ratings, highlighting the best and worst hospitals in the nation. Several medical residency programs even require their trainees to develop and execute a quality improvement project in order to complete their clinical training.

Yet, this widespread interest in quality and quality improvement is largely a modern phenomenon. Until recently, determinations of quality in medical care were left to professional judgment.[3] Individual physicians provided whatever care they felt was necessary and sufficient to treat patients' underlying medical conditions. Success was characterized by recovery from illness--unexpected success was often viewed as miraculous--while adverse outcomes were chalked up to "bad luck" or "sicker patients." Rarely, if ever, did external bodies, such as

hospitals, professional societies, or payers (e.g., public and private insurance companies), play a role in measuring health outcomes or directing the care of patients within their purview. Quality was simply left to the experts: physicians.

More recently, however, the assumption that clinicians always delivery the highest possible quality of care has been replaced with calls to measure and report patient-level health outcomes. By the mid-1980s, insurers, who had become used to paying physicians based solely on services rendered ("fee-for-service"), began to seek some justification for rapidly increasing medical costs. Quite simply, payers--and the patients they represented--demanded to know what they were getting for their money. In conjunction with improvements in the science of quality measurement and the increasing interconnectedness of the modern medical practice,[4] pressure from insurers has altered the quality landscape so significantly and so rapidly that many physicians now feel that the burden of proof rests on them to justify the treatments they prescribe or the procedures they perform.

In response to this external pressure, many hospitals have actually begun to flip the narrative and embrace quality as a core component of their business plan. Quality or quality improvement is often listed as a strategic goal of the organization and featured prominently in its mission statement. Many hospitals, especially academic medical centers, employ a Chief Quality Officer with similar levels of power and responsibility to other members of senior leadership. Some larger medical centers even have separate quality officers or quality committees for each clinical department. For hospitals, providing better care is more than an admirable pursuit; it is also a potential competitive advantage that must be capitalized upon if the hospital is to maintain its profitability and market share.[5] Under this vision of quality, improvement serves as necessary component of business development, like raising capital or attracting a competent labor force. As

businesses, modern hospitals--even modern non-profit hospitals--find themselves in a constant search for strategies and tools to help them improve care, attract new customers, and maintain or increase their bottom line.

The problem with this approach is that we as a healthcare system cannot reliably predict which strategies and tools will be most effective at improving clinical quality. Several notable success stories--a 25% reduction in preventable deaths at the nation's largest non-profit healthcare delivery system[6] or a 63% reduction in central line-associated blood stream infections at intensive care units (ICUs) across the country[7]--highlight the promise of certain well-designed and well-implemented quality interventions. However, an equal or larger number of failures to make, sustain, and translate improvements have forced us to question whether our current set of quality tools actually meets our quality needs.

For example, implementation of a 19-item Surgical Safety Checklist developed by the World Health Organization resulted in significant reductions to both peri-operative mortality and inpatient complications during an initial study of eight hospitals in eight cities around the world.[8] Yet, a subsequent, well-run implementation study of the same checklist at 101 hospitals in Ontario, Canada found no difference in either clinical outcome.[9] Clearly, either something about the checklist or its implementation was not replicated in Ontario or the quality problems that Ontario hospitals faced were not the same as those faced by the eight hospitals in the original study. Whatever the underlying reason, the failure of the Surgical Safety Checklist reflects a larger problem in quality improvement: even when hospitals know what to do differently, they often have difficulty figuring out how to make it happen. Reviewing guidelines does not seem to lead to changes in clinical practices and changes in clinical practices do not seem to lead to better patient outcomes.

This stagnation has contributed to a general disillusionment with the pace of quality improvement and a growing internal division between the individuals who actually deliver care and those who are tasked with trying to improve it.[10] Physicians are wary of spending time on what seem to be inconsequential, top-down quality efforts while administrators and senior leaders grow frustrated with physicians' apparent unwillingness to change. In an attempt to break this division and to accelerate improvement activities, some experts have suggested a new strategy: the formation of quality improvement collaboratives (QICs), which typically consist of groups of hospitals, often within the same geographic region, that work collectively to improve the quality of care for a single clinical condition.

QICs offer a slightly different vision of quality improvement than clinical registries or even single-institution quality improvement programs. Part data repository, part intervention incubator, QICs are design to provide the same level of data quality as national registries, but with the added benefit of actually using those data to develop shared efforts for implementing clinical best practices. As multi-institutional groups with a common goal, QICs promote the sharing of ideas between hospitals so that centers that perform well in one area can help others in the group and, in turn, receive help with topics on which they may still need to improve. This inter-organizational sharing of ideas also allows hospitals to overcome the "paradigm paralysis" that may have prevented them from recognizing their own deficiencies or adopting new improvement techniques.[11] With their larger sample size and the help of experts in quality methodology, QICs offer a level of methodological rigor that typically surpasses what any individual hospital might be able to provide. This external support insures that participants will learn from the data they generate during the collaborative process and be able to adapt any findings from other hospitals to the specifics of their local environments. Finally, many QICs are

intentionally rooted within a community, which allows hospitals to both broaden their perspective on what determine patient health (e.g., by focusing on social determinants) and to shift their focus from caring for individual patients to improving the quality of care for entire patient populations.

Yet, as hospitals are increasingly signing on to this new model of quality improvement several important questions remain unanswered. First, do QICs actually improve the quality of clinical care and, if so, how do they do it? Second, does participation in a QIC improve quality more than would be expected from single-institution quality improvement efforts? Put another way, do hospitals enjoy any marginal benefits from QIC participation? And, finally, what is the proper place for QICs among the larger armamentarium of quality improvement tools?

Attempting to answer these three central questions is the focus of my dissertation. As a surgeon involved in the formation of a trauma-system-wide QIC, I will, when possible, focus on the evidence with respect to QICs in surgery. However, as many of the larger issues with QICs--*if* they work, *how* they work, and *when* they work--remain universal, I would imagine that the majority of my results can be generalized to any group of hospitals working together to improve quality in a single clinical area.

What is quality?

Before I can feasibly hope to evaluate the effectiveness of QICs or any other method of improving quality, we need to understand what quality actually is as it applies to healthcare. Hundreds of definitions have been published since the earliest debates on healthcare quality and it is well beyond the scope of this dissertation to summarize the entire literature base. Instead, I will focus on three central components to quality that will be necessary to keep in mind as I

discuss the formation of QICs, how they contribute to quality improvement, and how we might assess their unique contributions to the field.

Number 1. Quality is determined, at least in part, by what we as physicians do and do not do for and to our patients. Quality can be influenced--for better and worse--by our actions. As clinicians, our decisions have consequences: adverse drug reactions, post-operative complications, missed opportunities to intervene. Do we make the correct diagnosis in a timely manner? Do we prescribe the right medication or perform the right operation? This interface between what, how, and when we *should* do to and for our patients and what, how, and when we *actually* do to and for them is at the crux of every definition of quality.

Avedis Donabedian, one of the fathers of the medical quality movement, defined quality as "that kind of care which is expected to maximize an inclusive measure of patient welfare, after one has taken into account the balance of expected gains and losses that attend the process of care in all its parts."[12(pp5-6)] Donabedian divided his definition of quality into the technical (i.e., what is done to and for the patient) and the interpersonal (i.e., how those services are delivered). A physician may choose the right medication for a given condition, but fail to explain to her patient why she is prescribing the medication or even what condition she has diagnosed. Conversely, a physician may recommend an operation when one is not indicated, but do so after taking into account her patient's goals of care. Neither situation fully lives up to the ideals of the medical profession or the capacity of the physician to care for her patient; both technical and interpersonal goals must be met to achieve high-quality care.

In fact, others have proposed a definition that explicitly takes this interaction into account. Brook and Williams proposed a similar model for quality, but with three variables: Quality of Health Care = Technical Care + Art of Care + (Technical Care) x (Art of Care) +

ε.[13(p134)] Beyond reminding us that every measurement has some component of error, Brook and Williams's model explicitly includes the interaction of technical and non-technical care, suggesting that each component on its own is of less importance without its complement.

How does this affect our understanding QICs? First, in order to improve patient-level outcomes, QICs must actually change how hospitals and the physicians who work in them take care of patients. This may occur either through a change in the processes of care (e.g., using a new, evidence-based antibiotic regimen) or a change in the structure of care that changes care processes (e.g., a new, multi-disciplinary ICU team that audits whether physicians wash their hands before entering patient rooms). Second, QICs can improve quality either by addressing the technical aspects of care or by improving the non-technical, interpersonal aspects of care delivery. In fact, both Donabedian's and Brook and Williams's models argue that addressing both aspects is necessary to truly achieve high-quality care. Therefore, in evaluating QICs and their success, we must look for patient-level changes across a wide spectrum of measurable outcomes.

Number 2. Quality varies in ways that do not reflect differences in disease severity or patient mix. Put another way, we as physicians cannot simply blame our patients for their poor outcomes; our patients are not as different as we may think. While many have contributed to this idea of "unjustified" variations in care, much of the movement, particularly the idea of unjustified geographic variation, is attributed to John Wennberg and the development of the *Dartmouth Atlas*. In 1973, Wennberg and Gittelsohn published a landmark article documenting differences in "bed and manpower use, expenditures, and utilization" across Vermont's 13 hospital service areas.[14(p1102)] The differences they found--from three-fold differences in the rates of appendectomy, mastectomy, and cholecystectomy to an over 10-fold difference in the rate of

tonsillectomy--did not appear to be due to disease prevalence or sociodemographics. Instead, Wennberg suggested that "factors intrinsic to the operation of the health care system," namely differences in provider and consumer behavior, were responsible for regional differences in healthcare utilization and spending.[14(p1105)]

Although the idea that non-clinical factors such as payment structure or facility ownership may drive differences in clinical practice sparked an entire movement in health services research, simply observing differences in outcome--even ones that do not appear to be justified by differences in disease severity--is insufficient to distinguish high-quality from low-quality care. Other types of analyses, such as process benchmarking (i.e., comparing how different hospitals or physicians perform the same clinical function) or the application of appropriateness criteria, are needed to pinpoint *where* technical and interpersonal care have met expectations and where they fall short. For example, in one of the most-cited articles on medical quality, McGlynn and colleagues applied 439 validated quality indicators for 30 acute and chronic conditions to a sample of nearly 7,000 patients through telephone interviews and medical record review. Overall, participants received only 55% of recommended care processes. However, while they did find significant and expected underuse of care--46% of participants did not receive at least one recommended care item--the authors also discovered that 11% of participants actually received care that was unnecessary and potentially harmful.[15(p2641)]

How should we use this information in evaluating the role QICs may play in improving patient-level health outcomes? First, to the extent that QICs contribute to documenting and explaining variations in care, they may help to identify areas where quality can and should be improved. Although, in some instances, single-institution data may be sufficient for recognizing signs of poor quality (e.g., by looking at inter-*physician* variation or by measuring levels of

guideline-discordant care), QICs may still fill a void by providing larger, more robust datasets that are capable of identifying inter-*hospital* variation in processes or outcomes for the same clinical condition. Second, if documenting variation hints at but does not localize or explain poor quality, then QICs must contribute more than clinical registries do; that is, they must provide more than risk-adjusted data reports. Inter-hospital comparisons of the type of care provided, how that care is delivered, and whether that care is appropriate are needed, if we are to view QICs as a success.

Number 3. There is no uniform way of defining or measuring quality; in fact, its definition often depends on the perspective of the interested party. While healthcare quality may not suffer from the same level of ambiguity that Justice Stewart famously faced in defining pornography (paraphrased: "I can't define it. But I know it when I see it."), it may come close. Without a shared definition of quality, what one observer deems high quality may be simultaneously--and rightly--viewed as average or even low quality by another. Therefore, to correctly interpret measurements of quality, we have to know exactly what is being measured, the strengths and weaknesses of the chosen metric, and the perspective of the person or group performing the evaluation.

A recent article in *Health Affairs* on hospital quality metrics helps to demonstrate this important point. Austin and colleagues compared how 844 hospitals performed across four rating systems used by well-known consumer groups: *U.S. News*, HealthGrades, the Leapfrog Group, and *Consumer Reports*. Not a single hospital in their analysis was rated as a "high performer" across all four systems nor was one consistently rated as a "low performer" across the three systems that included such a rating (HealthGrades only lists top hospitals).[16] More remarkably, only 10% of hospitals rated as a top performer by one system were also among the high

performers in any of the other systems and 27 hospitals were simultaneously rated as a high performer by one system and a low performer by another.[16(p427)]

Two important factors contribute to this apparent contradiction and both must be kept in mind when attempting to evaluate the impact of QICs. First, quality is multifocal--at the very least there is a technical and an interpersonal dimension--and different rating systems may focus on different aspects of quality. *Consumer Reports* and the Leapfrog Group attempt to measure patient safety while HealthGrades focuses on patient outcomes, specifically complication and mortality rates; *U.S. News* uses a more nebulous definition to identify the "best medical centers for the most difficult patients."[16(p424)] While there is as certain overlap between safety and outcomes (e.g., poor safety measures probably contribute to more complications and deaths), it is conceivable that a hospital could have reliable procedures to prevent "never events" like a retained foreign object or a wrong site surgery, but still have surgeons who lack important elements of technical quality.

Second, components of quality and quality measurement depend on perspective. While some measures are universal (lower complications rates, for example, would be viewed as higher quality by all parties), others remain hotly debated. Investment in technology, a structural measure that would improve a hospital's Leapfrog or *U.S. News* rating, may also be welcomed by patients as a proxy for high-quality care, which may, in turn, be reflected in higher measures of patient satisfaction. However, new technology often drives up the cost of care more than its quality, a result that may be viewed negatively by insurers or health policy experts. A thorough evaluation of QICs must, therefore, be performed from multiple perspectives: the patient seeking care, the hospital participating in the collaborative, the sponsoring agency, if one exists, and even

the individual teams of people from each participating hospital (e.g., physicians, nurses, administrators, and hospital leadership).

Measurement, feedback, and the movement toward QICs

Working from this concept of quality--one that requires an examination of what we do to patients and how we do it, one that focuses on identifying and eliminating unjustified variation in care, and one in which quality depends on perspective--it is not hard to see why some experts have become disenchanted with quality efforts that have focused primarily (and almost exclusively) on developing national clinical registries. Nor is it hard to see why QICs, which offer both quality measurement and the structured exchange of ideas on clinical practice, are seen by many as the next step in the evolution of modern quality management.

The majority of quality improvement manuals begin with the same important concept: the need for measurement. Before you can determine what to improve and where to direct your resources, these texts argue that you must first have good data. Data can objectively identify areas of poor performance; data can quantify differences in output or outcome between two productive units (i.e., physicians, wards, hospitals); and, most importantly, data can be tracked over time to help determine if changes that are made actually result in improvement.[17]

But generating and analyzing data is only a one step in the larger process of quality improvement. In *Curing Health Care*, Berwick and colleagues describe their method of process improvement as, "observe the world, understand the variability, formulate hypotheses, and then act on those hypotheses."[18 p. xv] Measuring and comparing hospital performance (i.e., outcome benchmarking) can teach hospitals where they stand on the quality curve and even allow them to track their performance over time as they experiment with different care models. But

measurement alone cannot help hospitals identify what they need to do to improve. Paul Plsek, another father of the medical quality movement, noted that, "organizations and individuals may know that their performance is significantly better or worse than that of their peers, but they may not know why."[11(p87)] In fact, our obsession with measurement may actually prevent us from making the changes we need to make in order to improve. In the preface to the paperback edition of their book (published thirteen years after the original hardback), Berwick and colleagues reflect upon the lessons they learned in putting their theories into practice. Lesson 1 includes a warning that, while important, teams may "unconsciously use [quality measurement] as a way to delay or avoid the discomfort of taking action, but improvement cannot occur without action."[18 p. xvi] Quality experts universally believe that measurement must be a part of the solution. Yet, many fear we have become stuck perfecting our measurement techniques rather than using the tools we have to improve clinical outcomes.

Few examples offer a better synopsis of the promises and perils of quality measurement than the National Surgical Quality Improvement Program (NSQIP). In the early 1990s, the Department of Veterans Affairs (VA) was concerned by anecdotal reports of variation in clinical outcomes for patients having surgery in its national system of roughly 150 medical centers. To get a better sense of what was actually going on, the VA collected pre-operative, intra-operative, and post-operative data (up to 30 days) on nearly 84,000 patients undergoing non-cardiac surgery at 44 of its medical centers between October 1991 and December 1993.[19] After adjusting for 128 patient-level variables, Khuri and colleagues were able to rank hospitals from 1 to 44 based upon their peri-operative morbidity and mortality rates. This approach--first correct for patient-level characteristics and then use hospital-level outcomes as measures of quality--has since formed the basis for hundreds of publications on hospital benchmarking, including the official launch of the

VA NSQIP program in 1998 based on data from 123 medical centers and 418,000 patients over a 6-year period (October 1991-September 1997).[20] For the first time, the VA was able to objectively identify 11 high-performing and 13 low-performing centers based purely on their risk-adjusted clinical outcomes. Combined with the success of similar programs for cardiac surgery in New York State,[21] the clinical registry quickly became the major focus of quality improvement in surgery.

By the early 2000s, NSQIP had entered the private market under the control of the American College of Surgeons, the largest constituent organization of practicing surgeons in the United States. Initial pilot studies found similar success at ranking hospitals to what had been described for VA centers[22] and comparisons with administrative coding--the only other available method of measuring hospital performance--consistently demonstrated better accuracy for the NSQIP data.[23] Buoyed by this success, NSQIP has become the most visible and most dominant surgical quality improvement program in the U.S. From three hospitals in its initial pilot, NSQIP has grown to over 500 participant hospitals today. Its data are routinely published in peer-reviewed journals and cited as the "gold standard" for surgical risk-adjustment. Its success has even spawned smaller, subject-specific programs, such as the Trauma Quality Improvement Program (TQIP) or NSQIP Pediatric, a database devoted to tracking pediatric surgical outcomes.

Yet, despite NSQIP's unparalleled success at *measuring* hospital quality, more controversy exists with respect to its ability to actually *improve* clinical outcomes. In fact, multiple evaluations of the program have produced disparate results. Two studies from the College itself using an interrupted time-series design have suggested improvement over time for the majority of NSQIP hospitals, with larger improvements being seen at hospitals that have spent more time in the program.[24,25] More recently, however, two papers adopting a difference-

in-differences approach found that, while NSQIP hospitals did improve over time, their rate of improvement was not significantly different from that of other hospitals not participating in the program.[26,27] In short, tangible improvements may have occurred, but these improvements did not necessary appear to be associated with hospitals' participation in the program. Other issues, such as selection bias (i.e., hospitals that would have improved on their own are the ones that choose to participate in the program) and the generalizability of findings to smaller, non-academic medical centers, also need to be answered before many will be prepared to ascribe the apparent clinical benefits to participation in the program. While even its toughest critics continue to commend NSQIP for encouraging the rigorous measurement of surgical outcomes, they have also increasingly begun to call for a platform that can make better use of these data to promote collaboration and encourage the development of interventions aimed at changing the way care is actually delivered.[28]

Enter quality improvement collaboratives. As I will demonstrate in depth in the next chapter, QICs were designed to push quality improvement beyond the capabilities of even the best clinical registries. By combining data sharing with multi-disciplinary discussions on effective clinical practices, QICs are not only able to explore the sources of variation, but also to institute changes designed to eliminate it. By focusing on a single clinical problem, QICs are narrow enough to digest, modify, and implement evidence-based practices across a wide group of hospitals. At the same time, their ability to generate and disseminate knowledge also allows them to accelerate the spread of innovation to hospitals across the country and the world.

The Los Angeles County Trauma Consortium

In the fall of 2013, I began my own experiment with QICs in helping to found the Los Angeles County Trauma Consortium (LACTC). Los Angeles County is among the most socially, ethnically, and geographically diverse places in the United States. At nearly 4,800 square miles, its borders contain mountains, forests, deserts, beaches, and cityscape--a total land area that would make it the 8th largest U.S. state. As of 2014, over 10 million people lived in the County, including some of the wealthiest and the poorest in the nation. Serving all of them is the Los Angeles County trauma system, a county-sponsored, public-private partnership that has been overseen by the Emergency Medical Services Agency (EMS) since 1983. In 2013 alone, the system fielded over 25,000 emergency department visits for trauma, making it among the most active trauma systems in the country.

In forming LACTC, our group hoped to build upon a previous partnership between the county's 14 trauma centers and EMS by adding content experts and health services researchers and shifting the focus of the group from administrative issues to quality improvement. Based on the interest of our partner hospitals, we began our collaboration with an analysis of traumatic brain injury (TBI), the most common indication for ambulance transport to a trauma center in the county and the primary source of nearly 1/3 of all injury-related deaths in the U.S.[29] Frustrated by reports of variable quality between participating hospitals, our group set about collecting data on all patients with severe TBI over a two-year period (2009-2010). In addition to the over 200 variables that are routinely abstracted from the medical record and entered into a secure electronic database as a part of state and county accreditation processes, trained trauma program managers captured injury- and treatment-specific information on each of the 844 patients identified during the study period. Other than cleaning and internal validation, these data

remained unanalyzed until the formation of LACTC and were provided to the research team in de-identified form for the purpose of system-wide quality improvement.

LACTC has several advantages, both as a method of improving care in Los Angeles County and as a case study in understanding the inner workings of QICs. As an improvement tool, LACTC is large enough to generate valid data and to observe variations in clinical practice, but centralized enough to allow for in-person collaboration and the implementation of system-wide clinical protocols. It is focused on a single clinical field (trauma surgery) and began by studying variation in a single clinical condition (TBI). It combines clinical experts in both trauma and neurosurgery with researchers experienced in quality improvement. Finally, by building on an established partnership, incorporating the strong relationships that already exist between trauma surgeons, and operating under the auspices of county government, LACTC has already become a validated source of information on quality improvement in trauma and has gained the trust of local leaders in the field. To the extent that "buy in" affects results, LACTC may be well on its way to achieving its goal of improving the care of traumatically-injured patients throughout Los Angeles County.[30]

Still in its formative phase, LACTC also provides a unique opportunity to examine and evaluate the QIC process. From the selection of a clinical target to its use of data and the organization of its membership, the choices LACTC has made as a collaborative will undoubtedly play a role in its success or failure. By scrutinizing these choices in the context of what is already known about TBI and what other QICs working in other clinical areas have done, I hope to determine not only if LACTC will achieve its goals, but how future QICs can be constructed to most effectively and efficiency improve health outcomes.

Specific aims

The purpose of this dissertation is to improve our understanding of QICs and the mechanisms through which they impact patient health. To do this, I have drawn both from the published literature on QICs and from original data generated by LACTC. Chapter 2 begins by summarizing the available literature on QICs and presenting notable examples of their successes (and failures) at improving patient-level health outcomes. Here, I am interested in answering the first part of my first central question: do QICs actually improve the quality of clinical care? Beginning with the first known examples of QICs (the Northern New England Cardiovascular Disease Study Group and the Vermont Oxford Network), this chapter catalogues the development of the QIC model through the Breakthrough Series program and ends with a summary of the QIC experience in surgery.

Chapter 3 provides a brief introduction to my original work and frames what I hope to learn about QIC from each original paper. Chapter 4 addresses both the second part of my first central question (how do QICs work?) and my second central question: does participation in a QIC improve quality more than would be expected from single-institution quality improvement efforts? To do this, I perform a meta-narrative review of the literature on QIC success and interview 13 key experts in the development of the QIC methodology. My goal here was to create a novel conceptual framework relating QIC participation to changes in the way quality improvement is performed at each participating hospital. This chapter also evaluates individual components of QICs, their contribution to QIC success, and the idea that certain hospitals or clinical topics might be more (or less) amenable to the QIC approach.

Chapters 5 and 6 present data from our own collaborative, the LACTC, to answer my final central question: what is the proper place for QICs among the larger armamentarium of

quality improvement tools? Each chapter explores a possible benefit of the QIC approach to hospitals' own quality improvement efforts. Chapter 5 focuses on the use of data as a guide for the improvement process and suggests that, by collecting new data and developing new statistical models, QICs might be able to act as "laboratories" for quality improvement in ways that larger, less flexible clinical registries cannot. Chapter 6 focuses on the quality improvement intervention itself and how QICs can evaluate whether national guidelines should be implemented at the local level. Here, I discovered that the QIC approach may actually prevent hospitals from wasting time and resources on trying to implement a set of best practices that does not meet their needs.

Although collectively my papers attempt to understand and evaluate the QIC approach as a whole, each original work also addresses a single specific aim and tests a single hypothesis:

Chapter 4's Specific Aim. To improve our understanding of QICs as quality improvement tools, particularly the mechanisms that might explain how, when, and where they are most effective.

> **Hypothesis**: QICs augment hospital-level quality improvement efforts by strengthening certain improvement domains; however, hospital-level success (and failure) is determined both by the QIC's ability to influence these domains and whether this support matches individual hospitals' quality needs (i.e., innovation-system fit).

Chapter 5's Specific Aim. To explore the ability of regional QICs to collect unique, valid, and condition-specific data and to use that data to refine and improve analytic techniques.

> **Hypothesis**: QICs can collect meaningful patient-level data that is not currently included in national clinical registries and can use that data to improve the way registries perform inter-hospital risk-adjustment.

Chapter 6's Specific Aim. To explore the importance of selecting a viable target for improvement and the role regional QICs can play in adapting national best practices to local environments.

Hypothesis: QICs can identify and evaluate potential targets for improvement and determine whether interventions developed at the national level can be successful within the context of regional care systems.

Finally, Chapter 7 summarizes the findings of my original work and then uses what I have learned to evaluate whether QICs do, in fact, impact quality, based on the framework that I presented earlier in this chapter. I conclude by further exploring the role that QICs can and should play as quality improvement tools and how they might interface with other types quality improvement programs. Specifically, should QICs operate at the national level or are they most effective when based in a single geographic region? If they remain regional, how should they relate to national clinical registries and to individual hospitals' own quality improvement efforts? Do QICs complement registries and single-institution quality improvement or supplant them?

Improving the quality of medical care remains one of, if not the major issue facing the future of American healthcare. Patients, insurers, and government regulators are no longer willing to give doctors the benefit of the doubt that the care they deliver is of the highest possible quality. Instead, they demand proof. While clinical registries can document quality, they have so far proved unreliable at actually changing clinical practices in ways that improve it. New strategies are clearly needed. Whether QICs can fill this void is the central focus of this dissertation.

Chapter 2 What are QICs and Do They Work? A Literature Review

While novel in their approach to quality improvement, QICs did not simply appear out of the blue in the mid-1990s. Nor does their unique methodology represent a departure from the principles of continuous quality improvement (CQI), QICs' substantive and temporal precursor among quality improvement methodologies. Instead, the development of QICs should be regarded as a step in the evolution of modern quality management theory: a continuation and extension of ideas that were developed in the 1930s, applied to medicine in the 1980s, and eventually adapted for multi-institutional teams in the mid-1990s. Given their evolutionary as opposed to revolutionary past, any proper evaluation of QICs must take into account the context of their development, specifically their reliance on a popular but somewhat unproven management philosophy in CQI.

Put another way, the fundamental question of whether QICs work, at least in terms of improving patient-level health outcomes, is really a series of questions. First, did CQI actually improve patient outcomes, and, if not, what limitations to the theory itself or to its implementation prevented it from doing so? Second, can the principles of CQI be adapted to multi-institutional collaboratives and, if so, can collaboratives based on these principles address the limitations CQI faced at the single-institution level? Finally, if QICs can, in fact, operationalize CQI at the multi-institutional level and address some of its prior limitations, have these QICs actually improved the quality of healthcare among their member hospitals?

To answer these questions, I performed a literature review addressing the following five topics:

1. The development and principles of modern quality management theory, specifically CQI

2. The effectiveness of CQI at improving patient-level health outcomes, both generally and in the field of surgery
3. Internal and external barriers to the success of CQI as it has been applied to the U.S. healthcare system
4. The development of a formal methodology for QICs and how this methodology was designed to address the barriers faced by CQI at the single-institution level
5. Published and unpublished evidence regarding the impact of QICs on patient-level health outcomes from early QICs, the Institute for Healthcare Improvement's Breakthrough Series, and QICs developed in the field of surgery

At the end of my review, I highlight several limitations to the current literature on QICs. Most prominently, there is no conceptual model to explain the pathways through which participation in a QIC affects patient health; instead, we are left to view QICs as a "black box" in which time, effort, and funding must be dumped with the outside hope that doing so will improve clinical outcomes.[31] Without a validated model, even the rare study that does demonstrate large-scale, empirical benefits for the QIC approach is confronted with skepticism and, at best, moderate disbelief. Correcting this deficiency by supplying a model that both explains the QIC methodology and accounts for variation among QICs and among hospitals within a single QIC is the purpose of Chapter 4.

The other major limitation to the literature--and the focus of Chapters 5, 6, and 7--is a discussion of QICs' role with respect to the larger world of quality improvement. By design, the majority of studies focus on the success or failure of a single QIC or, at most, several QICs using the same structure (e.g., the Chronic Care Model [CCM]), but not addressing the same clinical

topic (e.g., one QIC might use CCM work on improving diabetic outcomes while another works on outcomes for patients with congestive heart failure). The few studies that do attempt to generalize the results of QICs typically involve answering a larger clinical question (e.g., do pre-operative antibiotics decrease surgical site infection [SSI]?) rather than a larger methodological one (e.g., are QIC-based interventions as effective as single-institution interventions at increasing compliance with pre-operative antibiotics?). Drawing from my experience with our own QIC, I use Chapters 5 and 6 to address two ways in which the QIC approach might actually advance aspects of quality improvement more generally: 1) by improving methods for risk-adjustment and 2) by validating (or invalidating) proposed interventions prior to actually implementing them.

In addition to summarizing my findings, my final chapter discusses the larger relationship between QICs and national quality improvement efforts, another topic that is missing from the current literature. A recent upswing in participation for both QICs and national, registry-based quality improvement programs suggests that both approaches will continue to be a part of the quality improvement landscape for years to come. As such, not only will understanding the history, development, and prior experiences of QICs make future QICs more effective, it may also influence the way in which we approach quality improvement as a whole.

Building principles: continuous quality improvement

By the mid-to-late 1980s, many experts had become convinced that the emphasis on quality assurance in the late 1970s/early 1980s had not only been unsuccessful at improving care, but was quickly becoming harmful to establishing a future for quality improvement in medicine. From the beginning, quality assurance--a strategy for maintaining quality through surveillance

and regulation--had been viewed by clinicians as a top-down, administrative approach with administrative goals (e.g., saving money) rather than a genuine, ground-up effort to improve patient outcomes.[32] At its core, the idea of quality assurance seemed to tip the balance in the age-old debate between the art and science of medicine. Empirically, clinicians doubted that any systematic attempt to monitor and regulate treatment decisions would allow for enough flexibility to accommodate subtle but important clinical differences between patients. Viscerally, however, clinicians felt that their training and judgment were being called into question. If administrators could decide what treatments a patient did and did not need based on an algorithm, what point was there to the hours, days, and years they had spent actually *practicing* medicine? This disconnect between clinicians, who viewed quality as providing all the services they felt patients required, and administrators, who viewed quality as providing only the services they deemed appropriate, quickly led to an animosity that threatened the future of quality improvement in medicine.[33] Seeking an alternative approach, a new generation of quality experts began to look outside of medicine, hoping to improve quality not by designing and enforcing better rules, but by re-envisioning and re-engineering the entire healthcare delivery process.

This new vision of quality--alternatively termed continuous quality improvement (CQI) or total quality management--builds off of lessons from industrial engineering and the Japanese concept of *kaizen* ("continuous improvement"), a central factor in the ascension of Toyota Motor Company to its position as one of the most efficient and high-quality car manufacturers in the world. More than just a playbook of successful interventions, CQI combines a general philosophy on personnel management with a methodology for analyzing and learning from a wide range of clinical problems.[34] As a movement rather than single, prescriptive method, many descriptions and definitions of CQI have been published since its emergence in the late

1980s/early 1990s.[11,35-37] Rather than summarize this extensive literature base, I have chosen to focus on four key concepts that differentiate CQI from previous approaches to quality improvement.

Number 1. Processes affect outcomes not people. Among the first articles to apply industrial quality improvement techniques to medicine, Donald Berwick's "Continuous Improvement as an Ideal in Health Care" typifies the view of CQI proponents that outcomes--both good and bad--are the product of systems and processes rather than the individuals who make up these systems or carry out their processes. Berwick begins his article by comparing the hospital to a factory assembly line and asks readers to imagine two very different foremen.[38] The first monitors her employees with an eye toward identifying and punishing poor performers. Berwick calls this the Theory of Bad Apples: since poor quality is the result of poor execution, quality can only be improved by identifying and removing those employees who are unable to perform to the highest-possible standard. The second foreman adopts a fundamentally different approach, one Berwick calls the Theory of Continuous Improvement. Instead of looking for mistakes, this second foreman supports her workers, urging and helping each to improve so that the entire company may benefit from their newfound efficiency. Since even the most competent individuals are both guided and limited by the system in which they work, Berwick argues that quality can only truly be improved by "understanding and revising the production process on the basis of data about the processes themselves."[38(p54)] An often-quoted corollary to this principle is the idea that each individual system is perfectly designed to produce the results that it produces. If we want to improve the product, we have to re-design or re-engineer the system that created it.

Number 2. Trimming weeds is as important as fighting fires. While other approaches to quality--including quality assurance--focus on identifying and eliminating deviations from the

norm (e.g., the malfunctioning valve, the drunken assembly line worker), CQI fixates on measuring and improving the *stable* performance of the system. As with other quality improvement techniques, variation remains the enemy under CQI; unlike previous approaches, however, CQI targets an entirely different type variation. Kilo and colleagues describe this distinction as "common-cause variation," the domain of CQI, and "special-cause variation," the primary focus of quality monitoring techniques.[32] For example, an uncharacteristically high rate of ventilator-associated pneumonia in the ICU (e.g., 25% of intubated patients during one month compared to a baseline rate of 5%) reflects a special-cause problem, one that clinicians might solve by looking for deviations from the norm (e.g., a new class of ICU residents, a new protocol for cleaning ventilators between use). On the other hand, a baseline ventilator-associated pneumonia rate of 5% instead of 1% or even 0% reflects an entirely different, common-cause problem, produced by an entirely different set of clinical processes (e.g., the length of time that patients remain on the ventilator after a difficult operation).

Although potentially inefficient, this division between improving a system's static performance and its dynamic performance is not entirely surprising in medicine. Under a traditional quality management approach, stable but suboptimal performance rarely raises alarms with management who are often evaluated based on the quantity rather than the quality of a unit's output (e.g., bed turnover rather than the percentage of patients who contract pneumonia). To make matters worse, clinicians--who are also traditionally paid based on quantity rather than quality--are trained to think about patients as individuals and are often unfamiliar or uncomfortable with quality improvement at the system level.[10] As such, common-cause problems almost automatically become the responsibility of outsiders ("special projects teams" or "improvement staff") rather than remaining within the job description of the individuals who

actually deliver clinical care.[32(p5)] Without support from clinicians, even the most talented and motivated "improvement staff" struggle to impact provider behavior, leaving the system's baseline performance largely unchanged.

CQI offers a more-comprehensive, systems-based approach. Rather than reacting to problems as they emerge, quality efforts under CQI are both proactive and multimodal: *quality planning* builds capacity, *quality control* addresses current quality problems, and *quality improvement* streamlines care processes by empowering the entire organization to refine the clinical production line.[18(pp42-43)] Under the CQI approach, organizations do not simply wait to "put out" large quality fires; instead, they work continuously to identify and correct even small imperfections ("trimming the weeds") that might negatively impact quality.

Number 3. Follow the data. Although decidedly less exciting than Woodward and Bernstein's more famous catchphrase, "follow the data" has, in many ways, become the guiding mantra of the CQI movement. Rather than relying on individuals' well-meaning but often inaccurate quality assessments, proponents of CQI turn to data as the only unbiased judge of performance. Data can identify high-performing hospitals (e.g., those with low SSI rates or high patient satisfaction). Data can document how care is being delivered (e.g., how often the correct skin prep is used). Data can even help identify targets for improvement by comparing how high- and low-performing hospitals treat similar types of patients. Yet, for all of its reliance on data, CQI differs from other statistics-based quality techniques in two important respects: 1) the type of information it considers important (i.e., what constitutes "data") and 2) how that information is used to improve performance.

Prior to CQI, someone interested in hospital quality might have looked for a number--typically the rate of an objective outcome like mortality--and then compared that number either

among competing hospitals or within the same hospital over time. As I discussed in the previous chapter, this approach may, under the right circumstances, provide a comparative measure of performance, but it cannot explain *why* organizations differ in their performance or *how* to improve low-performing outliers. To actually understand these more nuanced but more important differences, CQI imported a different type of measurement (process benchmarking) and a different type of data (process data) from industrial quality management. Check sheets, swim lanes, fishbone diagrams, key informant interviews, and surveys quickly became the preferred instruments of quality improvement rather than complex but less descriptive outcome metrics.[39] On the analytic side, flow diagrams and simple graphs, which could easily translate findings so that workers could visualize change over time, took the place of computational statistical methods and hypothesis testing.[40]

CQI's belief in finding the right data to guide improvement--combined with its consumer-first mentality--also drove proponents toward involving patients and their families in the process of quality assessment. In their seminal article on CQI, Laffel and Blumenthal propose a unique and radical definition of quality improvement: "the continuous effort by all members of an organization to meet the needs and expectations of [patients and other] customer[s]."[41(p2870)] While this definition highlights nearly all of the principles of CQI, it is particularly notable for placing the patient, rather than the healthcare provider, as the final authority on healthcare quality. This idea that hospitals should seek to maximize both the product (i.e., health) and the experience of its consumers helped to put patient satisfaction on the map as a viable and important health outcome.

In addition to its unique vision of what constitutes data, CQI also adopted a new outlook on the role of data in quality improvement. Berwick describes this shift as moving from

"measurement for judgment" to "measurement for improvement."[42] Rather than focusing on data as a report card, CQI tries to use data descriptively (to determine what is actually going on), supportively (to celebrate high-performers), and prospectively (to help formulate solutions for future quality problems).[43] There are two main reasons for this shift from a punitive to a practical approach to data. First, using data as a report card harkens back to the Theory of Bad Apples where results are based solely on people instead of the systems and processes that influence their performance. In addition to being inaccurate, this outlook imposes a static view of quality that demotivates the entire workforce: since high performers will always be high performers, they have no reason to think about improvement and, since low performers will always be low performers, they have no hope of benefitting from improvement.

Second, many experts began to view the traditional tools of experimental science (e.g. randomization) as out of place when applied to quality improvement.[44] In eliminating bias, experimental isolation breaks the connection between quality improvement interventions (e.g., a nurse-based peri-operative checklist) and the context on which they often rely for success (a team of motivated operating room nurses). Because of this, not only may randomized controlled trial (RCT) evidence in quality improvement be difficult to come by, it may also produce spurious results and limit what can be learned from important failures. Instead, CQI's approach to data is practical: rather than focusing on *proving* a problem exists or that one solution is statistically better than another, we should "[collect] only the data [we] need"[40(p205)] to *suggest* that change is necessary or that a certain process will improve care and then to actually *test* these ideas with rapid-cycle improvement techniques. As I will discuss later in my section on QIC methodology, these rapid-cycle improvement programs allow organizations to learn from their own improvement efforts and build momentum for change over time.

Number 4. Sustainable change requires sustained commitment to change. CQI's final key concept involves applying its continuous and dynamic vision of quality to the organization itself. Some have referred to this idea as making the conversion from a standard organization to a "learning organization": one that promotes the acquisition and use of new knowledge, empowers its workforce to make changes,[41] and internalizes the drive for improvement instead of allowing its actions to be driven by external bodies (e.g., regulations, billing, mergers).[45]

In order to make this conversion, CQI urges organizations to transform their institutional culture by flattening the traditional hierarchy and encouraging the use of multi-disciplinary teams. Logistically, this transformation requires three phases of organizational change: 1) front-line providers (i.e., physicians, nurses, physical therapists, case managers, pharmacists, care partners, and even service staff) need to be re-invested in the quality improvement process; 2) senior leadership needs to demonstrate its support for quality improvement by providing both time and resources; and 3) front-line providers and senior leadership need to work together toward the same, mutually-agreed-upon goals. As Batalden and Buchanan put it: "everyone in the organization is involved in improving quality because everything can be improved."[46(p140)]

While all three phases are essential, the first component--empowering front-line providers--is often viewed as the most radical and the most difficult to accomplish. Berwick describes this process as re-establishing "respect for the health care worker."[38(p55)] Although it may appear to some as superficial, the decision to emphasize front-line providers throughout the quality improvement process is by no means empty flattery; it is both a practical and necessary step in selecting the best intervention and insuring its proper implementation. Proponents of CQI quickly realized that, unlike typical members of the C-suite, who either have no clinical training

or are far-enough-removed from their clinical practice to have lost touch, front-line providers understand what is actually driving gaps in care. In the Michigan Keystone Project, for example, it was front-line providers who suggested that the behavioral norm of treating central line infections as "inevitable" might be as important a contributor to breaks in sterile technique as issues with provider knowledge or the availability of sterile instruments.[47] This insight allowed the intervention team to reframe central line infections as a social problem with a commensurate set of social solutions (e.g., combining storytelling and community responsibility with data on infection rates). Moreover, since front-line providers are also often called upon to actually deliver many of the components of quality interventions, gaining their buy-in up front is essential to maintaining the fidelity of the intervention during its implementation. Operating room nurses who do not believe in safety checklists or do not believe that their role in administering safety checklists will be respected by surgeons are both less likely to implement the necessary process changes and less likely to do so with the energy and confidence needed to change surgeon behavior.

But, empowering front-line providers alone or even creating separate streams of influence for front-line providers and senior management still falls short of CQI's vision of the learning organization. W.E. Deming, one of the fathers of industrial quality management, included as the first of Fourteen Points on total quality management: "create consistency of purpose toward improvement of product and service."[36(p689)] Others have referred to this idea as engendering a "shared authority for quality" throughout the organization. True, senior leadership and front-line providers play different roles in executing this shared vision of quality, but both groups must work together at every stage in the process--from identifying the problem to planning and implementing solutions--to ensure that their shared vision becomes a reality.

With these four basic principles, CQI directly addresses all three aspects of quality that I presented in the previous chapter. First, by focusing on processes instead of people, CQI forces clinicians to confront what they do and do not do to and for their patients. Since both the decisions (i.e., technical quality) and the way these decisions are made and communicated (i.e., the art of care) contribute to the clinical outcome, both are appropriate targets for improvement under a CQI framework. Second, CQI explicitly focuses on identifying and eliminating variation in care and even provides tools to aid in this effort. CQI's more expansive vision of what constitutes data and how that data should be used allows organizations to not only quantify variation, but to examine and eliminate its sources by better understanding care processes. Finally, CQI embraces multiple new perspectives on quality and quality improvement. On the demand side, patients and their families are no longer viewed as passive recipients of care; instead, their satisfaction is an important determinant of quality that must be measured and improved under CQI. On the supply side, since every step in the healthcare production line--from the CEO to the environmental services staff that clean patients' rooms--can affect clinical outcomes, every member of the organization must contribute to a shared vision and a shared commitment to quality. Because of this shared responsibility, understanding how each group thinks about quality and allowing each group to participate in improving it is central to success under CQI.

Evaluating CQI's impact on health

For all of its novel methods and beliefs, CQI's impact on healthcare quality in the U.S. remains unclear. Many have credited the movement with fundamentally changing the way we view quality and quality improvement; others consider the movement largely unsuccessful at

translating its novel vision of quality into actual changes to patient care. In reality, both groups are probably correct, as I will demonstrate in the following section. Part of this disagreement over CQI's impact on health may simply be definitional: CQI can refer both to a general philosophy on quality and quality improvement as well as to interventions based upon this philosophy. As both are relevant to the development and success of QICs--the ideas behind collaborative quality improvement stem from the CQI philosophy while the actual activities carried out by QICs typically involve CQI-inspired interventions--I decided to review articles addressing both aspects of CQI.

As a management philosophy, the impact of CQI is most apparent in the way quality is discussed, measured, and intervened upon in the current healthcare environment. In their qualitative review of CQI, Blumenthal and Kilo focus on three areas of success, all of which involve CQI's influence on thought and practice rather than on actual patient outcomes.[34] First, CQI succeeded in transforming the language of quality improvement from punitive to positive. While this change in mindset occurred largely at the level of individual physicians, Blumenthal and Kilo note that many regulatory organizations, including the Joint Commission, have publically admitted that their previous definition of quality as compliance with a given set of standards was too modest a goal. In fact, many of these organizations actually changed the way they evaluate hospitals in response to CQI to better match a new vision of quality as the continuous examination and refinement of clinical processes. The Joint Commission's current mission statement exemplifies this transformation: "to *continuously* improve health care for the public, *in collaboration with other stakeholders*, by evaluating health care organizations and *inspiring them* to excel in providing safe and effective care of the highest quality and value" (emphasis mine).[48]

Second, CQI successfully validated the consumer as an important authority on quality and heightened the public interest in measuring and improving patient satisfaction. Today, hospitals compete to be viewed as the most "patient-centered," Medicare documents patient satisfaction as a quality metric, and funders of research, including the newly-formed Patient-Centered Outcomes Research Institute, put a premium on understanding how to make patients more satisfied with their healthcare experiences. Finally, Blumenthal and Kilo credit CQI with broadening interest in quality improvement from a select group of experts to the majority of hospitals and healthcare organizations across the U.S. In a 1993 national survey of 3,303 hospitals, 69% reported that they had implemented a CQI-based policy change, with 75% of these having done so in the previous two years.[49] In response to CQI, organizations with no previous experience in quality improvement began to experiment with their own improvement programs. Even clinicians, who had, until CQI, largely considered quality and quality improvement outside of their job description, felt empowered to experiment with the ways in which they delivered care in the hope of achieving better patient outcomes. In short, CQI transformed healthcare quality from a social experiment to a social movement, effectively and definitively bringing the issue into the public eye.

Yet, beyond its general influence on the field, establishing the impact of either CQI as a management philosophy or of CQI-based interventions on actual patient health remains a challenge.[50,51] Again, part of this difficulty may be methodological. In their commentary on the use of RCTs to evaluate the benefits of CQI, Samsa and Matchar highlight three important limitations.[36] First, the success of CQI may vary significantly based on participants' "buy in," making the choice of study population a potential source of bias. Testing a CQI-based intervention among a group of unmotivated participants may doom the program to failure before

it has even begun. At the same time, testing the same program among a group of highly-motivated participants might result in benefits that are unrelated to the intervention itself. Second, even the most effective CQI-based interventions may take years to affect patient outcomes due to the type of changes involved (e.g., improving operating room culture) and the number of affected parties (surgeons, nurses, techs, etcetera). Third, because of this protracted timeline, extensive resources are often required to properly document changes in process and outcome over time. In an era of diminishing funds for research--much less program evaluation--these types of studies rarely occur. Without them, qualitative studies on CQI may actually suffer from different type of publication bias: programs that are actually on track to make changes over time get evaluated too early, before measurably positive findings have been generated.

Despite these limitations, several studies have attempted to summarize the impact of CQI-based interventions on health and healthcare. Among the most-cited articles on the topic, Shortell and colleagues performed a systematic review of CQI-based interventions that targeted either overuse, underuse, or misuse of clinical services.[35] The authors identified 42 single-site and 13 multi-site studies, 11 of which (9 single-site and 2 multi-site) focused on surgical procedures. Overall, the results were favorable, however, none of the 3 RCTs (2 single-site and 1 multi-site) found a clinical benefit for CQI and there was considerable variability in target (the vast majority of studies addressed misuse), study design (most did not include a control group), and intervention parameters (most involved either provider training or education). The authors conclude that "although there are 'pockets of improvement,' no evidence has yet emerged of an organization-wide impact on quality."[35(p609)]

As a part of their previously-mentioned narrative review of CQI, Blumenthal and Kilo also conducted 19 semi-structured interviews with experts on quality improvement in an attempt

to identify specific examples of success or failure.[34] Aside from its general influence on the field, the authors conclude that there are no shining examples of CQI's success and that the leaders they interviewed could not produce data to support the use of CQI-based interventions to improve clinical outcomes. The authors do note, however, that the majority of CQI-based interventions they reviewed focused on improving administrative outcomes (e.g., lower costs, shorter length of stay) rather than patient health.

Barsness and colleagues' previously mentioned survey of 3,303 hospitals also compared hospitals that had performed at least one CQI-based improvement intervention with those that had not. Hospitals that had used CQI were, on average, more satisfied with their quality improvement efforts, reported more input from leadership, had higher perceived productivity and profitability, and even demonstrated statistically significant differences in cost savings related to improvement efforts .[49] Yet, despite these potential organization benefits, the authors found no differences in the perceived impact on patient outcomes between CQI and non-CQI hospitals. A similar study of 61 hospitals also found that hospitals participating in a CQI-based improvement program reported lower hospital charges and shorter lengths of stay for six clinical conditions compared to hospitals that did not participate; however, there was no difference between the hospitals in clinical outcomes.[52]

As a contrast, D'Andreamatteo and colleagues performed a systematic review to evaluate the use of Lean, a related CQI-based improvement technique, in healthcare. The majority of included studies involved Lean-inspired interventions in a hospital setting and reported positive impacts on both productivity and cost effectiveness.[53] A smaller group of studies attempted to evaluate Lean as a more general approach for promoting organizational development rather than as a part of a specific clinical intervention; again, the results were largely positive, particularly

for measures of teamwork and safety culture. In fact, no included study identified a negative effect of Lean.

Two systematic reviews specifically address the impact of CQI-based interventions in surgery. Nicolay and colleagues identified 34 articles reporting on the use of industrial quality management techniques in surgery, including nine specifically addressing CQI. Results varied in magnitude based on the targets of the intervention, however, all studies found a positive impact for CQI.[54] Of note, the majority of interventions covered by this review had clinical targets (e.g., rates of compliance with pre-operative antibiotics or β-blocker use, pain, unplanned extubation), although a few studies focused primarily on administrative measures (e.g., median start time delay in the operating room, cost per case).

In a more recent systematic review, Mason and colleagues identified 23 articles addressing Lean, Six Sigma (another related CQI-based improvement tool), or Lean Six Sigma interventions in surgery.[55] Several included studies demonstrated an impact on patient outcomes (e.g., reduced post-operative complications, reduced nosocomial infection rates, improved glycemic control) with one study even showing a significant reduction in the 30-day mortality rate for patients undergoing an operation to repair a fractured femur after the beginning of a Lean-based intervention that focused on multi-disciplinary care teams (11.7% pre-intervention vs. 6.7% post-intervention, $p<0.05$). Again, none of the included studies reported worse outcomes after the implementation of an intervention based on any of the three quality improvement methodologies. Unlike the studies identified by Nicolay and colleagues, however, the majority of studies included in Mason and colleagues' review focused on administrative outcomes, namely markers of pre-operative and intra-operative efficiency.

Limitations to CQI theory and barriers to success

Taken as a whole, the literature on CQI's success is decidedly mixed. CQI as a management philosophy appears to have almost certainly influenced the way clinicians and quality experts think and talk about quality improvement. Yet, almost no data exist to substantiate CQI's impact on actual patient health. In fact, the existence of a few individual success stories against the larger backdrop of anemic results appears to suggest that the success of CQI may depend more on how its principles are implemented than on their inherent value to patient health. As Shortell and colleagues reflect, "(1) the concept is great, (2) design is important, but (3) implementation appears to be everything."[56(p5)] If this is the case, then treating the American experience with CQI as a failure of implementation rather than a failure of concept may help us to understand the subsequent development of QICs and why experts believed that these multi-institutional groups might help hospitals overcome some of the barriers they faced in trying to roll out CQI.

In their 1995 review article, O'Brien and colleagues present a framework for understanding the variable implementation of CQI across hospitals based on four separate but highly synergistic dimensions: cultural, technical, structural, and strategic.[57] The cultural dimension refers primarily to organizational norms, particularly whether an organization's employees feel empowered to participate in the quality improvement process. (Think Toyota and the development of the *andon* cord.) Organizations that perform well in this dimension typically have forward-thinking leaders who are willing to try new approaches, learn from their employees, and publically demonstrate their commitment to the quality improvement process. The technical dimension refers to whether organizations train their employees in quality improvement techniques (e.g., swim lanes, fishbone diagrams) and whether they provide

sufficient data and analytic support for employees to actually take advantage of these techniques by making and evaluating changes in clinical practice. The structural dimension refers to the way hospital personnel are organized: do hospitals leverage their corporate structure efficiently? Do they promote the use of multi-disciplinary teams? Do they incorporate any pre-existing quality entities (e.g., quality assurance managers or reporting requirements) into their CQI implementation?[56] Finally, the strategic dimension refers to how well quality improvement is aligned with the organization's underlying (and collectively-held) strategic aims: is quality improvement a central part of the organization's business plan or is it tangential (or even contrary) to what employees view as the ultimate goal of their work?

Extending this hospital-level framework to the CQI movement as a whole allows us to understand its collective implementation, how it faired in each dimension, and whether there were particular barriers to its success. Culturally, CQI appears to have captured the minds, but not the hearts of the majority of healthcare providers, particularly physicians. Multiple evaluations of the movement suggest that physicians largely behaved as tentative partners in CQI-based interventions and remained skeptical that efforts were driven by a genuine desire to improve care rather than to save money.[33,34] Moreover, physicians, who were, for the most part, used to working independently, had difficulty submitting to what often felt like external, administrative initiatives with no chance of producing meaningful clinical changes. In fact, Blumenthal and Kilo, D'Andreamatteo and colleagues, and Mason and colleagues each commented on the lack of clinical targets addressed by the CQI-interventions they summarized.[34,53,55] Without physicians who felt empowered to participate, CQI efforts quickly found themselves in a worsening cycle of detachment, disillusionment, and eventual failure: physicians, fearing that their ideas would not be listened to, stopped volunteering to lead quality

interventions; quality staff, feeling no pull from physicians, began to focus more on administrative topics; and physicians, whose fears had now been validated, responded by refusing to participate at all.[33]

In the technical dimension, CQI generally scores well. The practical demands of the carrying out CQI--namely employee training and data infrastructure--prompted many hospitals to completely reform their human development and technology systems. Yet, despite these steps toward becoming more modern and adaptive healthcare organizations, many experts feel that hospitals misapplied these new tools during the early rollout of CQI. Rather than focusing on the larger idea of creating continuous and flexible improvement programs, early CQI programs frequently attempted to apply the same quality intervention and the same system of measurement to all quality problems and all clinical settings. As a result, many hospital were unable to progress beyond small administrative projects, which largely turned-off clinicians, who had difficulty distinguishing this formulaic version of CQI from their prior (largely negative) experiences with quality assurance.[33] This perception of CQI as quality assurance with another name was not aided by the fact that many organizations simply repurposed employees who had previously run quality assurance programs as the leaders of these new, CQI-based interventions without any additional training or support.

Other critics also note that, while hospitals may have provided initial funding for training and data management, many significantly reduced their support over time. For the most part, hospitals simply underestimated how resource-intensive CQI-interventions would be and how important continuous reinvestment was for maintaining momentum for change.[34] In an era of rising healthcare costs and a growing concern about the sustainability of the system, many

executives felt that the level of investment that would be required to do CQI well was simply too large to have to wait years or even decades before seeing its clinical and financial windfalls.

Despite inspiring significant technical changes, CQI's largest impact on the U.S. healthcare system may actually have been structural. A simple comparison of hospital "org charts" from before and after the CQI movement seems to support this claim. Modern hospitals include a multitude of quality officers, multi-disciplinary teams, and other groups tasked with measuring and improving clinical processes. (UCLA Health, for example, has separate institutes for innovation, value, and performance improvement in addition to a Chief Quality Officer and an Executive Director for Quality.) However, as many of these newly created quality positions were immediately filled with administrators and business people rather than clinicians, certain structural changes may have actually worsened the growing cultural divide between clinical and non-clinical staff. The few shining example of hospitals that succeeded using CQI undoubtedly represent those that were able to bridge this divide.

Despite mixed results in the other three dimensions, CQI's ultimate failure lay in its inability to influence the long-term strategic aims of American hospitals. Even if CQI did lead hospitals to make a series of cultural, technical, and structural changes, without genuine and unmistakable support from senior leadership, the majority of these changes would have appeared to employees as little more than outward responses to the growing popularity of the CQI movement. Put another way, most hospitals merely went through the motions of CQI, adopting its practices but not its core principles. In fact, many of the difficulties that CQI faced in the other dimensions--physicians refusing to adopt a new culture, the inflexible application of quality tools, the creation of quality officers who could not or would not interface with clinical staff--suggest that employees reacted to this lack of support by taking a cautious approach to

implementation, as if they were not truly convinced that their bosses actually believed in what they were asking their employees to do. Without a shared belief in the value of CQI, hospitals did not internalize its lessons and quality improvement continued to be viewed as something that was done at certain times by certain employees rather than something that needed to be done by all employees if an organization was to achieve its strategic goals.

What should we conclude about CQI based on its performance across these four dimensions? First, the fact that some changes were made and that some hospitals had success seems to substantiate the idea that industrial management techniques can work in healthcare. However, a second, and perhaps more important, conclusion is that CQI cannot work without a supportive environment. Context is everything; unless interventions are directed at the right target, with the right implementation, in the right settings, and under the right leadership then they stand no chance of actually uprooting and reforming prevailing practice. Shortell and colleagues discuss this inherent fragility of CQI with colorful analogy:

> CQI may be thought of as a beautiful rose growing in an unruly garden filled with weeds…For the CQI rose to flourish, it must be carefully cultivated in a rich soil bed (e.g., a receptive organization), given constant attention (e.g., sustained leadership), assured of appropriate amounts of light (e.g., training and support) and water (i.e., measurement and data systems), and protected from damaging pests (e.g., overly burdensome regulation and parochial views).[35(p605)]

Clearly, American hospitals in the 1980s did not possess the contextual pre-requisites for CQI to be a success nor could the movement establish enough footing among physicians and senior leaders to generate these supportive factors on its own. Faced with this reality, experts began to shift their approach to quality improvement from trying to find a single, perfect model ("the CQI rose") to creating an environment in which this model could freely grow. It was this new challenge--developing a platform that could more efficiently translate knowledge, empower

providers, and align the perspectives of senior leadership with those of front-line providers--that led to QICs.

Toward a formal QIC methodology: the Breakthrough Series

Frustrated by the limited success of CQI, a group of experts in the early 1990s--led primarily by the newly founded Institute for Healthcare Improvement (IHI)--developed a new strategy for accelerating the pace of quality improvement in medicine. Despite its previous success at generating clinical knowledge and teaching the basics of CQI, the IHI had become concerned that a growing interest in quality improvement was not being translated into better clinical outcomes. Instead, a substantial gap was developing between generally agreed upon best practices and actual patterns of care delivery; a gap Charles Kilo, one of the leaders of the IHI's new approach, referred to as the division "between what we know and what we do."[58(p1)]

Rather than a lack of knowledge or tools, leaders at the IHI believed that this gap primarily reflected an inability to translate findings from one context to another, specifically from one hospital to another. Even though hospitals faced many of the same quality problems (e.g., hospital-acquired infection, unplanned readmission, long wait times to see a primary care physician), the existing, single-institution model of quality improvement forced each to design and test its own series of interventions. This process seemed relatively inefficient, especially since the IHI knew that solutions to many of the most common clinical problems had already been developed. Moreover, hospitals--even extremely motivated hospitals--could only design and test a limited number of interventions on their own. That meant that until hospitals were able to identify a workable solution, they would continue to deliver low quality care and might even harm patients during the improvement process. Instead, the IHI believed that, by sharing data

and experiences across institutions, hospitals could build upon what was already known and cut down on the time between identifying a quality problem and implementing its solution.

Under this premise, IHI co-founder Paul Batalden began to develop a structured program for producing "breakthrough changes" in healthcare by combining two separate learning systems: technical knowledge (what he refers to as "know what") and experiential learning (what he calls "know how").[59] To do this, several changes would need to be made to previous IHI programs. First, clinicians could not simply listen to lectures on quality improvement or take part in programs designed by IHI staff. Instead, they needed to apply what they were learning directly to their own clinical environments. Second, institutions needed to be represented by multidisciplinary teams from different organizational microsystems rather than by any single individual. Batalden felt that aligning these groups from the beginning of the program would force them to develop a single quality improvement strategy rather than separate initiatives within each sector (i.e., a nursing intervention, a physician intervention, etcetera). Third, and most importantly, individual providers or hospitals were abandoned as the object of improvement efforts. Instead, *groups* of providers and hospitals would be brought together to address a single, shared clinical problem, with the assistance of clinical experts and experts in quality improvement methodology. As Donald Berwick put it, the goal was to get hospitals to "compete against disease, not against each other."[42(p842)]

And so--on a napkin--the Breakthrough Series (BTS) was born.

Although the BTS model remains the most formalized and, in many ways, the gold standard, multiple frameworks for what constitutes a QIC exist in the literature. In his initial writings on collaborative quality improvement, Plsek described QICs as multiple organizations that participated in benchmarking efforts, identified and shared best practices, and then measured

improvement over time.[11] Kilo included any groups that collectively identified a topic, consolidated the relevant literature, and then used a formalized method to spread new knowledge across organizations in the group.[60,61] Øvretveit and colleagues proposed a similar definition to Kilo's, but required groups to use experts, develop measurable targets, and have both in-person and virtual group meetings.[62] Schouten and colleagues drew from all three of these previous definitions in developing their five essential features of QICs: 1) a specified topic, 2) clinical experts and experts in quality improvement, 3) multi-disciplinary teams from multiple sites willing to improve and share, 4) a formalized model for collecting data and testing changes, and 5) a series of structured activities in a given time frame.[63] More recently, Nadeem and colleagues expanded upon Schouten and colleagues' definition by clarifying that the topic must involve specific provider practices or patient outcomes and that the model must involve rapid-cycle quality improvement with "small tests of change."[64(p359)] Since the BTS fulfills all five of these definitions and remains the reference point for much of the literature on collaborative quality improvement, I will use it here to describe the logistics of the QIC process and to contrast this process with that of single-institution CQI.

Logistically, the BTS process can be divided into two groups of people and five distinct time periods. A Planning Group--typically a combination of 1) *subject matter experts*, who bring a knowledge of the clinical topic, 2) *application experts*, who are experienced in quality methodology, and 3) *success stories*, individual clinicians or hospitals that have demonstrated breakthrough performance on their own--is tasked with providing a framework for collaborative activities and administering the group's logistics (e.g., when the groups meet, where they meet, etcetera).[58] Under this Planning Group are teams from each of the participating organizations. Although the composition of these teams can vary slightly based on the topic and the selected

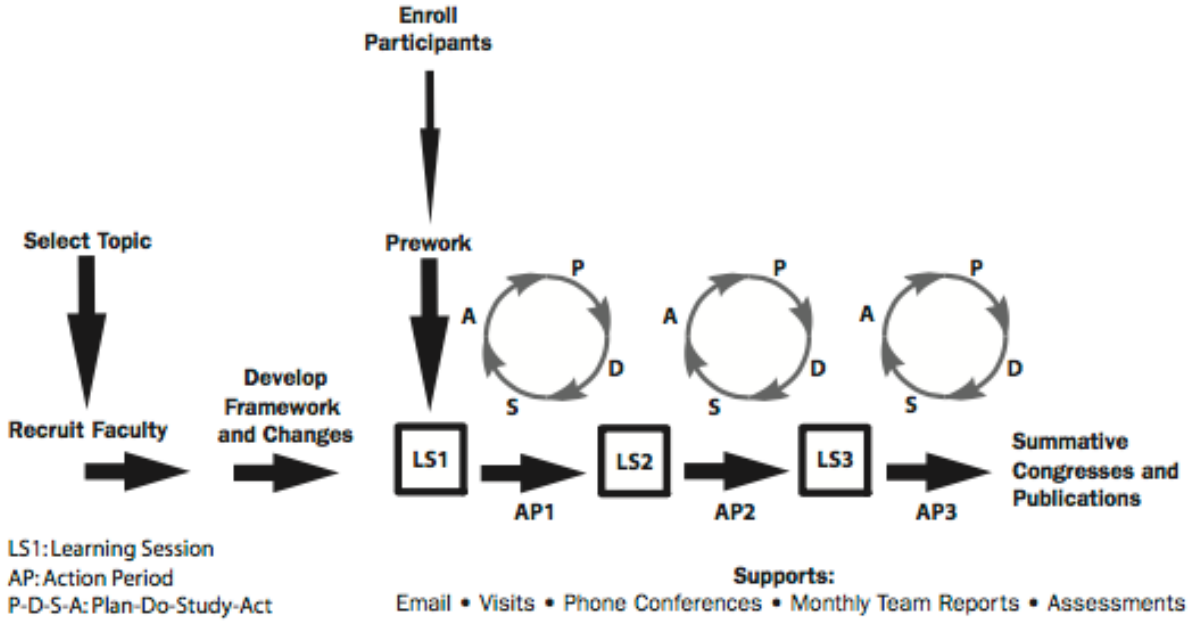
activities, each organization is tasked with identifying 1) a *technical expert* on the topic, typically a physician champion in the field, 2) a *day-to-day leader*, someone who understands the details of the system under study, and 3) a *system leader*, someone from senior leadership with "enough clout…to institute change."[60(p5)] Each of these team members plays a different role in the collaborative process and is responsible for performing a different set of tasks over the five phases of the project: pre-recruitment, pre-meeting, meeting, post-meeting, and summation (see Figure 2.1).

The pre-recruitment phase (the first vertical arrow and the first two horizontal arrows at the left end of the figure) is accomplished by Planning Group and involves three major goals: 1) the selection of a target topic, 2) the recruitment of any additional experts (e.g., more specific subject matter experts), and 3) the development of an appropriate set of interventions, measurement strategies, and attainable goals.[58]

This final component, called the Change Package, forms the core of the intervention and, in many ways, the core of the collaborative itself. Kilo refers to developing the Change Package as establishing the "theory for improvement"[60(p6)]: the process of identifying the root causes of poor performance and pairing these root causes with interventions aimed at correcting them. Others at IHI divide this theory for improvement into two: "content theory," the likelihood that the Change Package will improve outcomes if properly implemented, and "execution theory," the likelihood that the Change Package will be properly implemented.[65] A well-designed Change Package addresses both components.

For example, re-dosing prophylactic antibiotics at the appropriate interval during long operations has been suggested as a means of reducing SSI rates.[66] If anesthesiologists are not re-dosing antibiotics because they do not know that they need to, then an educational campaign may

Figure 2.1: Breakthrough Series Model (taken from Institute for Healthcare Improvement, 2003)[49]



be effective; however, if anesthesiologists know they need to re-dose, but are unable to leave the operating room in order to retrieve the needed antibiotics, then a different, systems-based intervention may be more appropriate. Here, the choice of Change Package addresses what the IHI refers to as "content theory": only by understanding the root causes of the problem can the right intervention be selected.

Beyond understanding the underlying mechanisms, a well-designed Change Package must clearly define the quality problem, adapt any general concepts to the local context,[40] and establish a reasonable set of expectations for the implementation process. These factors represent the "execution theory" behind the particular quality intervention. For example, rather than a general mission statement (e.g., "to reduce caesarean section rates"), a well-designed Change Package might read: "caesarean section rates in the Unites States can be safely reduced to less than 15 percent from the national average of 24 percent while maintaining or improving maternal and fetal outcomes."[60(p6)] By combining the appropriate content theory with the proper execution

theory, Change Packages allow QICs to lay out a series of specific changes (e.g., prevent admission for false labor, manage pain to help women tolerate labor, etcetera[67]) that can be applied and tested throughout the collaborative process.

The pre-meeting period (the two vertical arrows over "LS 1" in Figure 2.1) includes the enrollment of collaborative participants, the formation of teams, and the "pre-work" each team must perform in order to participate in the program. The importance of pre-work is two-fold. First, requiring organizations to commit time and energy (in addition to financial resources) prepares participants for the demands of the collaborative process. Common tasks in the pre-work assignment include forming the improvement team and establishing clear, quantitative, and organization-wide goals that can be used as a roadmap for success.[60] Second, the results of the pre-work process can help to improve the initial Change Package by crowdsourcing solutions to the clinical problem at hand. Since the entire QIC process is based on the idea that individual organizations have already discovered unique and successful methods for improving care, it is important that the Planning Group allows ideas from the pre-work period to influence and improve upon what it has already learned from its initial literature review and its discussions with clinical experts. Moreover, since many organizations in the early stages of improvement do not have a clear understanding of how they care for their own patients much less how their care practices differ from those at other hospitals (e.g., what percentage of patients having surgery get pre-operative antibiotics? Which antibiotics? At what time? Who is responsible for making sure this is done?), forcing organizations to review their own practices, their own data, and their own goals for participation also ensures that their time spent in the meeting sessions is as productive as possible.

The meeting period ("LS 1-3" in Figure 2.1) is perhaps the phase for which the BTS is best known. Over a period of 12 months, teams meet in person with the Planning Group for three separate two-day Learning Sessions. The first session is typically didactic: teams learn the basics of CQI from quality experts and the Planning Group presents its initial Change Package to the group as a whole.[58,60] As the collaborative begins to discuss the Change Package, attention is placed on defining terms and laying out a measurement strategy so that teams can present the status of their interventions when they return. In the second and third learning sessions, the roles are flipped.[65] Teams take turns presenting their data and sharing particular successes or failures. Because the entire collaborative is focused on the same clinical problem, presenting individual experiences offers organizations the opportunity to get feedback from other teams or to hear about different approaches that they may not have thought of or tried. Without this outside perspective, clinicians can become stuck trying to tweak their current practices instead of fundamentally reengineering them, a phenomenon author Joel Barker refers to as "paradigm paralysis." Without appropriate support, even if clinicians are exposed to a new approach, they may dismiss it as inapplicable to their current practice or try it, but soon revert to the old method with which they feel more comfortable.[11] By creating a protected, collaborative space for discussion, the BTS process encourages free exchange of ideas and generates a peer group to provide teams with the necessary support. Unlike single-institution quality improvement, QICs are based on solving a specific problem both as a group and for the group as a whole.

Although I have referred to the next period ("AP 1-3" in Figure 2.1 along with the associated "P-D-S-A" cycles) as "post-meeting" for simplicity, a more correct term might be "inter-meeting" since it is closely tied to the events of the Learning Sessions. This time is also commonly referred to as an Action Period because it involves teams taking elements of the

Change Package back to their local environment, studying the impact of these new ideas on their current practices, and then making changes to the protocols to better fit their needs. Unlike best practice guidelines, the BTS focuses less on getting teams to implement a fixed list of items and more on the iterative refinement of a set of ideas that have been developed as a group. There are two reasons for allowing and even encouraging organizations to influence the intervention itself. First, the success of quality improvement initiatives depends in large part on tailoring interventions to local context.[68] Actions that might bring about change in one institution (e.g., reorganizing clinical teams) might be met with skepticism or even worsen the problem in another. Since individual teams are experts in their own context, allowing organizations to choose which pieces of the Change Package they implement and how those pieces are applied provides the best opportunity to affect practices.

Second, encouraging teams to participate in defining the intervention creates ownership and increases the motivation with which the teams will work to ensure that changes are made. The alternative--forcing a foreign set of ideas on the organization--elicits what Pascale, Sternin, and Sternin call "the immune rejection response."[43(p112)] Since the team does not identify with elements to the intervention, they will not support them to others in the organization and the whole group may even band together in rejecting these new ideas. Under the BTS process, ideas that are brought back to the organization are already supported by local leaders, who act as liaisons between the QIC and the local environment.

Central to the activities of both the meeting phase and the post-meeting phase is the Model for Improvement, a series of quality improvement methods developed by the IHI in partnership with Associates in Process Improvement. Like the BTS as a whole, the Model should be thought of as a framework rather than a script. Under the Model, teams are asked to define the

goals of their project ("what are we trying to accomplish?"), to design a method for determining if they are meeting these goals ("how will we know that a change is an improvement?"), and to brainstorm potential interventions ("what change can we make that will result in improvement?").[17] After these questions have been answered, teams choose an intervention and test it in their local environment using Plan-Do-Study-Act (PDSA) cycles.

Also referred to as Shewhart cycles after their creator, engineer and father of statistical quality control, Walter Shewhart, PDSA cycles build on the Model's three questions by providing a framework for testing change. Unlike a small manufacturing business, medical systems are complex and often nonlinear, such that small changes in one area can result in large, somewhat unexpected changes in others.[69] By linking small, rapid-cycle changes to immediate reflection and rigorous outcomes tracking,[44] PDSA cycles offer a safe environment for teams to learn what works in their local context, to rapidly acquire technical skills in quality methodology,[41] and to build momentum for larger, system-wide reengineering projects.[40] Just as paradigm paralysis prevents organizations from considering new ideas, the static momentum of the status quo often prevents organizations from adopting them. Unlike traditional statistical methods whose large-scale, all-or-nothing data collection only reinforces this inaction, PDSA cycles allow teams to comfortably but rigorously make changes to clinical practice and to maintain a tension for change within the system such that early setbacks do not result in the total abandonment of quality efforts.

Finally, the summation phase ("Summative Congresses and Publications" at the right end of Figure 2.1) provides both an outlet for celebrating success and a structured process for disseminating knowledge outside of the current collaborative members. Each official IHI BTS project produced a document summarizing its work, many of which can still be found on the IHI

website (www.IHI.org). These documents include information about the choice of topic, definitions, measures, outcomes, and even a performance assessment for each organization that participated. High performers are typically highlighted in the executive summary by name while a detailed description of the Change Package and the individual intervention components is provided so that organizations interested in tackling the same quality problem could use the document as a guide. Again, this decision to publicize the details of the work is not surprising given the IHI's broader focus in developing the BTS framework ("to accelerate the rate of diffusion of existing science into clinical practice"[61(p384)]). Whether outside organizations have taken advantage of this information by directly implementing interventions from QICs of which they were not a part remains unknown. However, the IHI itself has built on its results; 10 of the first 26 BTS QICs were actually updates of previous BTS QICs and used the previous summary document as their initial Change Packages.[58]

In reviewing the BTS process, it is clear that the founders viewed the program as an extension of single-institution CQI methods rather than a new type of quality improvement activity. An early article by BTS founder Charles Kilo lists six fundamental principles that guided the development of the program:

1. A substantial gap exists between knowledge and practice in health care.
2. Broad variation in practice is pervasive
3. Examples of improved practices and outcomes exist, but they need to be described and disseminated to other organizations.
4. Collaboration between professionals working toward clear aims enables improvement.
5. Health care outcomes are the results of processes.

6. Understanding the science of rapid cycle improvement can accelerate demonstrable improvement. (taken from Kilo 1998[60])

From this list, the connection between BTS and CQI is unmistakable. The same focus on developing a systems approach, understanding baseline variation, analyzing data, and employing interdisciplinary teams that defined CQI remained central to the new program. In fact, several of the BTS principles appear to be drawn directly from these four fundamental CQI concepts. For example, "health care outcomes are the results of processes" reiterates CQI's focus on processes rather than people (CQI concept number 1) and "broad variation in practice is pervasive" echoes CQI's focus on baseline variation (CQI concept number 2). Principle 6 and principle 4 similarly tout the importance of data and teamwork, a direct linkage to CQI concepts number 3 and 4, respectively.

CQI and QICs differ, however, in two main respects. In a practical sense, QICs focus on implementing what is already known about a given clinical topic more generally (e.g., what is known about interventions to prevent surgical never events?) while CQI requires teams to generate new knowledge about either the topic, the intervention, or the clinical site before making changes (e.g., what has led to surgical never events and what might prevent them in this specific institution?). Even though the root causes to many common clinical problems overlap (e.g., poor communication can lead to both medication errors and wrong site surgery[70]), CQI does not allow hospitals to make assumptions about their quality problems and instead requires teams to explore and confirm their suspicions using data. While this process may occasionally result in a more accurate understanding of the problem--assumptions, even good ones, can at times be wrong--its length and rigor can also lead to frustration, loss of motivation, and even

withdrawal of support from senior leadership if certain deliverables cannot be achieved. In fact, one of the IHI's prime drivers in developing the BTS program was its own members' dissatisfaction with the slow pace and limited clinical impact of earlier IHI-sponsored quality improvement projects.[60] BTS QICs, on the other hand, last an average of only 17 months, only three of which are spent brainstorming and organizing prior to taking action.[62] In an era of shrinking budgets and increasing demands on time, the rapid-pace and greater efficiency QICs offer organizations an attractive alternative to the longer and more resource-intensive CQI process--so long as QICs are at least as effective as single-institution CQI.

The second major difference between QICs and single-institution CQI is the availability of external agents to help guide the improvement process. These agents--both experts and peers--affect hospitals in two main respects. First, listening to experts who have worked on similar clinical problems or with similar improvement teams can help hospitals generate new ideas and approaches. This guidance, combined with hearing how other teams from similar institutions have tried to solve the same clinical problem, forces hospitals to consider new perspectives and to break their own paradigm paralysis. Rather repeatedly trying to tweak their own quality programs, teams in QICs often discover and then implement new interventions that have either worked in other settings or on other topics (e.g., applying the ICU model to high utilizers of outpatient services in so-called "ambulatory intensive care units"[71]). Although QICs do not typically generate new *technical* knowledge, they do provide a platform for generating and transmitting *experiential* knowledge--a process that was largely absent from the quality improvement toolkit at the time.[59] In fact, many believe that learning what has worked and what has not worked for other organizations may actually be more helpful to hospitals than learning how to design their own programs independently.

Second, having data from multiple organizations allows QICs to directly compare each hospital's performance to others in the group. On a technical level, benchmarking allows hospitals to understand where they sit on the performance curve and allows the group as a whole to identify and reproduce the practices of high-performing hospitals--activities that would be impossible for individual hospitals acting alone. However, from an organizational management perspective, these same inter-hospital comparisons also allow QICs to create and maintain a "tension for change" across the entire group of participants. As I discussed earlier, keeping clinicians motivated during quality improvement programs can be difficult due to the slow pace of work and the often tangential connection to day-to-day clinical care.[34] However, when clinicians are placed in teams and supported by other like-minded individuals, they are more likely to become engaged and stay engaged throughout in the quality improvement process.[72] Some have suggested, more cynically, QICs increase motivation by increasing competitive pressure rather than improving support (i.e., no one wants to be to lowest-performing organization in the group or the one that cannot make the necessary changes). Either way, if structured, inter-organizational contact can make employees more motivated and more accountable for their performance, then QICs may offer hospitals a series of organizational benefits that are beyond what can be generated through even the best single-institution improvement programs.

Evaluating QICs' impact on patient-level health outcomes

As with CQI, the potential for collaboratives to improve clinical care is readily apparent. Yet, just as with CQI, I believe that the true impact of QICs must be judged by their results. I will, therefore, use this section to review the literature on whether QICs actually impact patient-

level health outcomes and, to the extent I can, whether they are more effective than single-institution efforts as a quality improvement tool. The literature on QICs can essentially be divided into two categories: 1) papers attempting to evaluate the impact of an individual QICs on clinical outcomes; and 2) papers documenting the formation, structure, and activities of individual QICs. While important for documenting the breadth of the QIC approach, this second category of studies provides little-to-no information on its effectiveness. Most employ a cross-sectional design, do not include a control group, and provide either qualitative data or data in a form that is not suited to meta-analysis. Unfortunately, these studies also represent the vast majority of the published literature on QICs. Therefore, while I will begin by discussing the few studies that attempt to measure the effectiveness of QICs in a formal manner, I will also supplement these data with studies that more informally present the trajectory of clinical quality before and after the formation of individual QICs.

Formal evaluations of QIC effectiveness

To date, two systematic reviews and one non-systematic review directly address QIC effectiveness. In their narrative review--the first on the topic--Øvretveit and colleagues present the results of two international research conferences on the QIC methodology. Although the focus of their article is "lessons learned" from QIC implementation, the authors do present the results of three successful BTS projects: a reduction in neonatal intensive care unit infection rates from 22% to 5% over 2 years, a reduction in caesarean section rates for 80% of hospital participants over 12 months, and a reduction in adverse drug events for 90% of hospital participants over 15 months.[62] The authors also describe in less detail two successful UK

collaboratives, a primary health collaborative that reduced delays between primary and secondary care and an orthopedic collaborative than decreased mean length of stay by 12%.[62]

Since summarizing the data on QIC is not the main focus of their review, Øvretveit and colleagues include a general discussion of QIC effectiveness rather than a meta-analysis of completed projects. Overall, the authors conclude that "quality collaboratives have had some success and that many teams and organisations taking part have benefited."[62(p350)] In particular, the authors note QICs' ability to "build interprofessional cooperation," improve relationships between colleagues, and "make significant clinical and organizational performance improvements more quickly than they might have done on their own."[62(p346)] Although anecdotal, this final statement is one of the few pieces of evidence directly comparing the QIC approach to single-institution quality efforts. Despite this generally positive outlook, the authors also present three potential areas of weakness for the QIC approach: 1) up to 30% of teams drop out of QICs before their conclusion, 2) only 30% of teams achieve "significant improvements," and 3) while there are some indications that outcome improvements can be sustained, there is "less evidence of continuous improvement or institutionalisation of the [collaborative] methods."[62(p349)] Taken together, these last three statements appear to suggest variability in results between organizations participating in the collaborative as well as questions about the sustainability of results after the formal collaboration has ended.

Despite the growing interest in QICs throughout the 1990s, the first systematic review of QIC effectiveness was not performed until 2008. Schouten and colleagues identified 72 articles that both met their definition of a QIC (see page 44) and included data on either care processes or clinical outcomes; 57 of these 72 articles involved QICs based on the BTS model.[63] Despite the

relatively large number of articles, most had serious limitations in design, including the absence of a control group (60 of the 72) or a baseline measurement of performance (number not stated).

Among the 12 higher-quality articles reporting on nine studies (two RCTs and seven pre-post, non-equivalent control group studies), there was conflicting evidence on QIC effectiveness. Both RCTs found little-to-no impact: one study that randomized pediatric practices in Detroit and Boston to either an intervention combining BTS methods with the Chronic Care Model or usual care found no differences in either process or outcome measures for children with asthma[73] while a study that randomized neonatal ICUs to a either a multifaceted QIC intervention or usual care found a significant improvement in two process measures (receiving surfactant at all in the delivery room and receiving the first dose within two hours of birth), but no differences in clinical outcome.[74]

However, among the seven controlled but non-randomized studies, six found more significant improvements in at least one metric for the sites that participated in a QIC than for the control (five of six BTS studies and the only non-BTS study). While the majority of studies focused on process rather than outcome measures, there were several notable examples of clinical improvements: a reduction in pain for patients at 21 nursing homes,[75] a reduction in infant mortality for a rural Alaskan medical center,[76] a reduction in infection rates for 10 neonatal ICUs,[77] improved hemoglobin A1c and cholesterol values for 698 patients with diabetes,[78] and increased peak flow rates, patient self-management, and quality of life for adults[79] and children with asthma.[80] Given the limited number of studies, the negative results from both RCTs, and the various methodological limitations, the authors conclude that "the evidence underlying quality improvement collaboratives is positive but still limited and that the effects cannot be predicted with great certainty."[63(p9)]

Nadeem and colleagues performed the only other systematic review on the topic in 2013 to both update the findings of Schouten and colleagues and to understand individual components and processes used by QICs, a topic that Schouten and colleagues were unable to address. The authors identified 12 new articles reporting on 11 experimental or quasi-experimental evaluations of QICs that had been published since the previous review (a thirteenth article, which updated a study already captured by Schouten and colleagues' review was also included).[64] The authors then combined these 12 articles with the 12 articles reporting on nine experimental or quasi-experimental studies from Schouten and colleagues to form an aggregate sample of 24 articles. Of the 24, 19 articles reported on provider-level processes measures, with 17 of these 19 (89%) finding positive or mixed positive results. Only 13 of the 24 included articles reported on patient-level outcomes; however, nine of these 13 (69%) found positive or mixed positive results.[64] Based on their update, the authors conclude that there is "further evidence that QICs can effect[sic] change at the provider level, particularly the process of care variables," but that "patient-level variables were assessed less often and the findings were decidedly more mixed."[64(p384)] In addition, despite identifying 14 separate QIC components and processes, the authors were unable to determine if any particular component or any particular QIC framework (e.g., a BTS versus a non-BTS approach) was associated with collaborative success.

For completeness, I performed my own search in October 2015 using the same search strategy and QIC definition as both Schouten and colleagues' (search date: June 2006) and Nadeem and colleagues' reviews (search date: April 2012; see Appendix 2.1 for search strategy). During the three and a half years since the last review, only one experimental or quasi-experimental study has been published on the topic. Power and colleagues reported the results from Stroke 90:10, a QIC based on the BTS model that randomized 24 National Health System

hospitals in northwest England to either the intervention or usual care. Among the 18 hospitals for which data were available, QIC participants demonstrated a 10.9% larger increase in compliance with one clinical bundle and a 11.2% larger increase in compliance with a second bundle relative to hospitals in the control group (p=0.025 for the first bundle, which focused on improving the timeliness of acute stroke care, and p=0.023 for the second bundle, which focused on post-acute rehabilitation).[81] Despite these promising findings, there was no difference in compliance between the groups for several of the most important process measures within each bundle (e.g. CT scan within 24 hours of hospitalization: -0.9% relative difference, p=0.799; occupational therapist assessment within 4 days of admission: +1.4% relative difference, p=0.789) and the study did not attempt to measure changes in clinical outcome.

Although slightly tangential to the larger question of whether QICs impact patient-level health outcomes, two pairs of articles report specifically on the cost-effectiveness of the QIC model. I decided to include them here because I believe they complement the limited data on changes in clinical outcomes and because they provide a direct comparison between the QIC-approach and the single-institution approach to quality improvement, albeit in monetary terms. Rogowski and colleagues published a companion article to their clinical report on the NIC/Q Project[77]--an early QIC operating within the Vermont Oxford Network (VON) that was included in both systematic reviews (discussed in detail on pages 68-70)--in which they analyze the financial aspects of collaborative participation. During the two-year study period, the ten intervention sites spent between $42,298 and $88,385 on time and travel resources, but yielded between $500,00 and $4.5 million in cost savings, based primarily on $10,932 reduction in the median treatment cost per low-birth weight infants (all amounts in 1996 dollars).[82] Although Rogowski and colleagues do not link them directly, the accompanying clinical report by Horbar

and colleagues[77] found a significant reduction in nosocomial infections--one of the clinical targets of the NIC/Q Project--among the intervention sites, which may explain the apparent cost savings. The median treatment cost per low-birth weight infants actually rose $12,249 in the control group during the same time period; there was no significant differences in treatment cost between intervention and control sites for infants with chronic lung disease, a separate target of the Project.

Researchers from the Dutch Institute of Healthcare Improvement similarly published a pair of articles from its QIC on diabetes management in 2010. The first, Schouten and colleagues (2010A)[83], reports clinical outcomes for patients seen at six of the eight sites participating in the QICs and for patients at well-matched controls. Although there was no difference between baseline and one-year after the QIC's end date, patients in the intervention group did have a modest but significant improvement over baseline in systolic blood pressure, mean high-density lipoprotein, and mean hemoglobin A1c levels at two years; however, there was no difference between these trends in the intervention and the control sites, suggesting that the QIC did not provide marginal benefit. However, in their companion article, Schouten and colleagues (2010 B)[84], the authors estimated that receiving care at a site participating in the QIC increased male patients' life expectancy by 0.97 years (0.44 quality-adjusted life years [QALYs]) and female patients' life expectancy by 0.76 years (0.37 QALYs). At €22 per patient to implement the program plus additional costs due to longer life expectancies and additional care, the QIC approach was calculated to have a €6,672/QALY incremental cost-effectiveness ratio for men and a €7,614/QALY incremental cost-effectiveness ratio for women, which is significantly below the Dutch standard at the time of €20,000/QALY. Therefore, while the clinical evidence suggested that sites participating in the QIC did not achieve significantly better outcomes than

matched controls, the cost-effective analysis concludes that "a QIC for diabetes care can substantially increase patients' lifelong health-related quality of life and life expectancy, even with modest results in the short run."[84(p890)]

Just as with CQI, when taken as a whole, the formal evidence on QIC effectiveness is limited and decidedly mixed. Of the five RCTs of QICs--the methodological "gold standard" for causality--only one produced positive results: a significant improvement in hemoglobin A1c, cholesterol, and compliance with foot and eye exams over 18 months for diabetic patients in clinics randomized to a BTS-based QIC compared to no change for clinics randomized to usual care.[85] The remaining RCTs predominantly showed no additional benefit associated with QIC participation, although the control groups varied from usual care to performance feedback to even a comparison of QIC approaches. As I discussed earlier in the chapter, some experts have begun to question or even dismiss evidence from RCTs in quality improvement since randomization inherently separates the intervention from important contextual factors that may drive its success.[44] As such, it is unclear what to make of the fact that QICs do not appear to perform well (or at least as well) when organizations participate based on randomization instead self-selection into the program.

Drawing solely from the quasi-experimental results, the evidence for QICs appears to be stronger, especially with regard to its impact of provider-level processes of care. The vast majority of studies (17 of 19) found improved compliance with target metrics (typically the delivery of evidence-based best practices) for providers participating in QICs compared to either no quality improvement intervention or one that did not involve structured sharing across multi-institutional groups. The breath of topics in which QICs have been successful--from pain control in nursing homes[75] to childhood development,[86] heart failure,[87,88] and organ donation[89]--also

appears to support the idea that the QIC methodology is not specific to a given clinical topic or setting, but can be used more generally as a quality improvement tool.

However, despite several suggestions of success, I believe that caution is still warranted before concluding that QICs are an effective quality improvement tool, much less a more effective tool than single-organization CQI. First, there remain several high-profile examples of QIC failure, most notably Landon and colleagues' study (included in both systematic reviews), which found that patients receiving care at clinics participating in a BTS-inspired QIC for HIV care did not have lower viral loads, higher rates of appropriate screening tests or prophylaxis, or better access to care compared to patients receiving care at clinics not participating in the QIC.[90] (Interestingly, viral load actually improved in both groups over the study period, but there was no significant difference in change over time between the groups [+11.0% of patients controlled in the intervention group vs. +5.4% of patients controlled in the control group, p=0.18].)

Second, even the small set of studies that does include a comparison group often does not match intervention hospitals to comparable sites or describe the type of quality improvement activities that take place in the control group. To illustrate this point, I abstracted all of the available information on the intervention and control groups from studies included in Schouten and colleagues' review (see Table 2.1). Other than the two RCTs, only one study explicitly mentions matching intervention and control sites and it does so based solely on relatively unspecific structural factors. (Another study vaguely mentions "comparability" based on the opinion of collaborative leaders). In addition, only one study explicitly reports what type of quality improvement activities are taking place at control sites. Without these pieces of information, it is difficult to truly compare QICs to hospitals attempting to solve similar quality problems using a single-institution approach. Add to this the issue of selection bias, the

Table 2.1: Descriptions of the Intervention and Control Groups from Schouten and colleagues' review

Study	Topic	Intervention		Control		Measures	Findings
		Subjects	QI Activities	Subjects	QI Activities		
Horbar et al. 2001[77] Rogowski et al. 2001[82]	Neonatal intensive care	All 10 NICUs in VON that participated in the NIC/Q project	Site specific implementation of "potentially better practices" for preventing nosocomial infection and chronic lung disease (number of specific changes not reported)	66 NICUs in VON that did not participate in the NIC/Q project	Not stated (other papers note that all VON hospitals participate in outcome benchmarking)	*Outcome*: coagulase-negative staph infection, need for oxygen supplementation at 36 weeks, death	2/3 significantly better 0/3 significantly worse
Benedetti et al. 2004[78]	Diabetes	All 11 primary care providers who attended 3 collaborative learning sessions	Implementation of 6 components of the CCM: clinical information systems, decision support, delivery system design, self-management, community, healthcare organization (number of specific changes not reported)	19 primary care physicians in the same group practice who did not participate in collaborative learning sessions	Not stated	*Process:* annual A1C, LDL, urine protein, eye exam, foot exam; proportion of patients taking aspirin *Outcome:* A1C<8.0%, A1C<9.5%, LDL<130, BP<130/85, BP<140/90, met self-management goal	7/12 significant better 0/12 significantly worse

Horbar et al. 2004[74]	Neonatal intensive care	57 NICUs in VON that were randomized to attend a workshop designed to teach the basics of collaborative quality improvement	Didactic sessions, facilitated site team exercises, multi-institutional group exercise designed to promote 4 key habits: change, evidence based practice, system thinking, and collaborative learning (number of specific changes not reported)	57 NICUs in VON that were randomized to control	Standard outcome benchmarking reports	*Process*: surfactant treatment in the delivery room, timing of first surfactant dose administered *Outcome*: death before discharge, pneumothorax	2/4 significantly better 0/4 significantly worse
Landon et al. 2004[90]	HIV	44 primary care clinics receiving funding under the CARE Act that participated in a BTS collaborative, agreed to participate in the study, and provide chart review data (of 62 collaborative members)	Site specific (43 change efforts per site on average) in 5 areas: antiretroviral therapy, screening, access to care, prevention, women's health	25 primary care clinics receiving funding under the CARE Act that were matched to intervention sites based on the type of site, location, number of locations delivering care, region, and number of patients with active HIV	Not stated	*Process*: MTB screening, flu shot, hepatitis C screening, Pap smear, receipt of PCP screening, receipt of HAART, had a visit in at least 3 or 4 quarters *Outcome*: viral load <400 copies/mL on last visit, proportion of patients with increased CD4 count	1/8 significantly better 0/8 significantly worse
Asch et al. 2005[87] Baker et al. 2005[88]	Congestive heart failure	4 organizations that participated in a BTS collaborative and agreed to participate in the evaluation (of 14 collaborative members)	Site specific (42 change efforts per site on average) in 6 domains: self-management support, delivery system design, decision support, information support community linkages, and health system support	4 sites that "had not participated in the collaborative but were otherwise comparable organizations with respect to structural characteristics reported by site leaders"	Not stated	*Process*: 21 indicators for screening, medication, follow-up, and counseling *Outcome*: BP<130/80, BP<140/90, INR 2-3, LDL<100	9/25 significantly better 0/25 significantly worse

Homer et al. 2005[73]	Childhood asthma	22 primary care practices that were randomized to participate in a learning collaborative for 12 months	Implementation of CCM, exposure to guidelines from the National Asthma Education and Prevention Program (number of specific changes not reported)	21 primary care practices that were randomized to control	Not stated	*Process*: written asthma plan in past 12 months, daily use of an inhaled steroid *Outcome*: daily use of a controller medication, any asthma attach in past 12 months, mean limitation from strenuous exercise, any asthma hospitalization in past 1 month, any asthma ED visit in past 12 month	0/7 significantly better 0/7 significantly worse
Mangione-Smith et al. 2005[80] Schonlau et al. 2005[79]	Childhood asthma	9 organizations that participated in the Improving Chronic Care BTS collaborative, provided care to children, and agreed to participate in the evaluation (of 24 collaborative members)	Not stated	4 internal comparison sites from these same 9 organizations	Not stated	*Process*: patient self-management (3 items), asthma knowledge, use of long-term controller medications *Outcome*: quality of life, impact on family functioning, satisfaction with care, acute care use, missed school days, parent missed work days	4/14 significantly better 0/14 significantly worse

A1c, hemoglobin A1c; BTS, Breakthrough Series; BP, blood pressure; CCM, Chronic Care Model; HAART, high-active anti-retroviral therapy; INR, international normalized ratio; LDL, low density lipoprotein; MTB, mycobacterium tuberculosis; NICU, neonatal intensive care unit; VON, Vermont Oxford Network
Horbar et al. 2001 and Rogowski et al. 2001 as well as Asch et al. 2005 and Baker et al. 2005 describe the same intervention and control

significant variation in success between collaboratives and even between sites within collaboratives, and the fact that mechanisms by which QICs affect patient-level health outcomes remain a "black box"[31] and it is not surprising that there is skepticism and continued debate over the usefulness of the methodology

Informal evaluations of QICs, summary reports, and descriptions of change over time

Although they offer less in the way of formal evaluation, reviewing several notable examples of QICs over time may help to illustrate the range of topics and implementation strategies used by QICs and to shed some further light--albeit mostly anecdotal--on the general strengths and weaknesses of the methodology. To do this, I have broken the remaining literature into three sections: 1) a narrative review of the published literature on the two QICs that formed prior to the BTS, 2) a narrative review of the initial BTS projects, and 3) a systematic review of QICs that have been founded in general surgery.

Early QICs. For a field that is typically associated with hierarchy and tradition (at least in the lay media), it is slightly surprising that the first documented example of a QIC took place in surgery. In 1987, all 18 cardiothoracic surgeons performing coronary artery bypass graft (CABG) surgery in Maine, New Hampshire, and Vermont joined together to form the Northern New England Cardiovascular Disease Study Group (NNECVDSG). As their first joint project, the group decided to create a prospective clinical registry of all patients undergoing isolated CABG surgery in the region to determine if hospitals varied in their risk-adjusted inpatient mortality rates. Their initial results, which included 3,055 patients having an operation between July 1987 and April 1989, suggested a two-fold difference in mortality among hospitals (3.1% to 6.3%) and an almost five-fold difference in mortality among surgeons (1.9% to 9.2%).[91]

In response to these findings, NNECVDSG began a three-pronged, two-year intervention in 1990 that drew heavily from CQI methodology, but incorporated a broader, more collective focus.[92] First, the group decided to provide feedback to surgeons and hospitals at least three times per year on their risk-adjusted performance so that each surgeon or hospital could see where they stood in the context of the group. Second, the group scheduled a two-day training conference for its hospital executives and two four-hour training sessions for all of its staff on quality improvement methodology. Finally--and most novel, even for today--the group began site visits to each of the five hospitals, which took place over a four-month period and included an industrial engineer in addition to visiting surgeons, surgical nurses, and administrators. Plsek refers to this practice as "process benchmarking": directly comparing hospitals not by their results ("outcome benchmarking"), but by *how* they provide care for patients with the same clinical condition.[40] Although the majority of modern QICs do not physically visit other hospitals in the group to learn from differences in practice patterns, the idea of each team openly sharing how it addresses the same clinical scenario and how it has overcome the same organizational barriers is central to the QIC concept.

In 1996, NNECVDSG published its results from 15,095 patients treated over six years (1987-1993). Compared to baseline rates (1987-1990), the group found a 24% reduction in risk-adjusted mortality during the post-intervention period (1991-1993; $p=0.001$).[92] This reduction in risk-adjusted mortality occurred in both men and women (19% [$p=0.01$] and 31% [$p<0.001$] respectively), but appeared to be isolated to urgent and emergent cases (25%, $p<0.001$) as opposed to elective ones (21%, $p=0.11$). Yet, despite the statistical significance of these changes, it is difficult to attribute this apparent drop in mortality to participation in the QIC without being able to compare these trends to those of a well-matched control group.

Three additional factors may help to put these findings into context. First, the authors also report that process benchmarking and re-engineering resulted in 18 specific changes in technical care, organization, and evaluation that "were [both] substantive and…temporally associated with the reduction in the mortality rate."[92] This temporal connection between process and outcome lends some support to the intervention (i.e., the QIC formation) being the source of lower mortality rates. Second, at 5.8%, the annualized nationwide reduction in mortality for CABG surgery during that time period based on Medicare data was substantially lower than the roughly 10% drop seen in Northern New England.[93] Although not a direct comparison, these secular trends do not appear to fully account for the observed mortality difference in the NNECVDSG. Finally, the contemporaneous (and unrelated) launch of New York State Cardiac Surgery Reporting System (CSRS), which mandated outcome reporting and limited CABG operations to high-volume hospitals, reported a 41% reduction in mortality during roughly the same time period (1989-1992).[21,93] Although, again, not a perfect comparison, this final piece of data somewhat qualifies the evidence in favor of QICs, at least in terms of their effectiveness relative to other types of quality improvement strategies. Still, NNECVDSG's ability to facilitate inter-hospital collaboration, to transform processes of care, and to contribute to improvement in patient-level health outcomes (if not to definitively affect them) validated the QIC model among quality experts and set the stage for future efforts, including the BTS.

The second major QIC operating prior to the BTS was the Vermont Oxford Network (VON), which was established in 1988 with the goal of "improving the effectiveness and efficiency of medical care for newborn infants and their families through a coordinated program of research, education, and quality improvement projects."[94(p350)] Since then, VON has continued as a non-profit corporation, funded by its members to act as a data repository (the VON

Database) and a coordinating center for research on neonatal outcomes. Beginning with 34 centers in 1989, the VON Database now receives data from over 1,000 centers worldwide and maintains a registry of over two million low- and very-low-birth-weight infants who were either born at a VON member hospital or transferred to one within 28 days.[95] As a part of its mission to improve the knowledge base surrounding neonatal care, VON also sponsors RCTs, outcomes-based research, and quality improvement research among its member hospitals, including several formalized QICs.

Two such QICs are worth further discussion: the Neonatal Intensive Care Collaborative Quality Project (NIC/Q) and the Evidence-based Quality Improvement Collaborative for Neonatology, really a series of QICs that take place every two to three years on different sub-topics within neonatology (the first meeting was confusingly stylized as NIC/Q 2000, although there was no formal connection to the first NIC/Q Project). Beginning in 1995, NIC/Q in many ways paralleled the BTS structure, but within the already formed partnership of VON. At the outset of the project, multidisciplinary teams--neonatologists, nurse mangers, administrators, and quality improvement coaches--from 10 neonatal ICUs met to review process and outcome data from the VON Database in order to choose a clinical target and to begin developing outcome measures. All 10 teams decided to work on nosocomial infection rates and developed a list of 17 "potentially better practices" that were then distributed to all participants.[77] While teams were asked to document which "potentially better practices" they put into place, each ICU was encouraged to choose only the most relevant changes instead of creating uniform treatment protocols. Over the course of three years (1995-1997), the 10 teams participated in four in-person group meetings, 14 site visits, 2 benchmarking visits to high-performing sites, and 64 conference calls. As I mentioned in the previous section, a clinical report from the project, which was

published in 2001 and included in both Schouten and colleagues' and Nadeem and colleagues' systematic reviews, found a significant reduction in the rate of nosocomial infections among participant hospitals, from 22.0% in 1994 to 12.3% in 1996. Of note, this trend represented a significantly larger reduction than was seen for the 65 VON-member hospitals that did not participate in the project (10.3% to 7.2%; p=0.001 for the comparison of trends between NICQ/Q and non NIC/Q VON hospitals).[77] A follow-up study, which was also included in both systematic reviews, randomized hospitals to either participation in a similar collaborative intervention or to usual care (all VON member hospitals had been receiving hospital-specific performance reports since 1990). Interestingly, this study found increased rates of compliance with two process measures (the percentage of low-birth weight infants receiving surfactant at all in the delivery room and the percentage of low-birth weight infants receiving the first dose of surfactant within two hours of birth) for hospitals participating in the QIC, but no difference in their rates of pneumothorax or mortality.[74] Taken together, these two studies suggest that self-selection of hospitals, the choice of clinical topic, or other unmeasured differences in hospitals may be key to QIC success.

In the NIC/Q 2000 project, teams from eight neonatal ICUs met in 1998 to review the available evidence on the treatment of bronchopulmonary dysplasia. The group then developed a list of nine "potentially better practices" for the condition and returned to their institutions to implement the changes they felt were most relevant. All eight sites went on to participate in the next two QICs in the series, NIC/Q 2002 and NIC/Q 2005, although the topics no longer focused on pulmonary care. A recent study by Payne and colleagues found that, despite not working on bronchopulmonary dysplasia after 2001, compliance with the "potentially better practices" developed as a part of the NIC/Q 2000 project at the eight original hospitals continued to

increase over time. In particular, hospitals reduced their use of delivery room intubation (70% vs. 52%), conventional ventilation (75% vs. 62%), and postnatal steroids (35% vs. 10%) while increasing their use of nasal continuous positive airway pressure (57% vs 78%, $p<0.001$ for all).[96] Trends in the original outcome measures over time were more mixed: there was no difference in rates of bronchopulmonary dysplasia-free survival (68% vs. 66%, $p=0.16$), although nosocomial infection rates decreased (18% vs. 15%, $p=0.045$) and rates of survival to discharged increased over time (90% vs. 93%, $p<0.001$). Without a comparison group, however, the true effect of QIC participation along with the importance of contextual elements, such as the competing demands on hospitals participating in multiple QICs, provider motivation, and specific QIC characteristics, cannot be definitively assessed. Still, it appears that at least some process changes generated by QICs may persist over time, at least in hospitals that are motivated enough to voluntarily participate in multiple quality improvement efforts.

Early Breakthrough Series QICs. In its 2003 white paper on the BTS methodology, the IHI provides a list of the 26 official BTS QICs it helped to facilitate between 1995 and 2001.[58] Sixteen of these 26 QICs were on unique clinical or administrative topics while 10 represent updates of previous QICs (e.g., "Reducing Cesarean Section Rates II (1997)" builds on the findings from "Reducing Cesarean Section Rates I (1995)" with a different group of hospitals). Of the 16 unique topics, I was able to find either a summary report or a published article on 10; I could find no reports or articles from "Improving Prescribing Practices (1997)," "Providing More Effective Care for Low Back Pain (1997)," "Improving Access and Efficiency in Clinical Office Practices (1998)," "Improving Service Quality (1998)," "Improving Safety in High-hazard Areas (2000)," and "Quantum Leaps in Patient Safety (2001)." Based on current

estimates, the IHI has sponsored over 50 QICs involving over 2,000 teams from 1,000 separate healthcare organizations.

In 1995, the IHI launched its first three BTS QICs: "Improving Asthma Care for Children and Adults," "Reducing Cesarean Section Rates," and "Reducing Delays and Waiting Times." The first project, "Improving Asthma Care for Children and Adults," involved 12 teams, focused on implementing clinical best practices for asthma, and had four major outcome goals: 1) reducing admissions, emergency department visits, and repeat hospitalizations; 2) improving functional status; 3) increasing the use of appropriate inhaled medications; and 4) reducing cost.[97] Interestingly, this first QIC included both clinical (e.g., improving functional status) and administrative outcomes (reducing costs). A summary report from the project includes examples of individual organizations meeting each goal, but does not provide a formal assessment of the project other than rating each institution on a scale from 1 to 5 based on its results, with 1 signifying no change and 5 signifying "outstanding, sustainable results."[98(p169)] For this project, three organizations were rated between 2-3, two were rated between 3-4, three were rated a 4, and the remaining four were rated between 4-5.[97(p177)] This summary score approach to evaluation was subsequently used by the IHI to rate the performance of teams for each subsequent BTS project.

The second BTS project, "Reducing Cesarean Section Rates," reviewed the literature on rising rates of cesarean section and developed 11 key change concepts (e.g., preventing cesareans for failed induction of labor, avoiding hospital admission for false labor). Twenty-eight teams then spent 12 months implementing these change concepts with the goal of reducing cesarean section rates by 30%. At the end of the study period, four teams (15%) achieved this goal while another 14 (50%) reduced their cesarean rate between 10-30%.[67] Overall, seven teams were rated

below a 3, ten were rated a 3, seven were rate between 3-4, one was rated a 4, and the final team was rated between 4-5.

"Reducing Delays and Waiting Times," the third BTS project, chose as its target a 50% reduction in delays and waiting times in four clinical areas within 12 months: the OR, the emergency department, clinics and physicians' offices, and general access to care. Individual metrics were developed for each area (e.g., time from triage to treatment room in the emergency department, median holding time to be admitted to a ward unit).[98] All but one of the 27 teams demonstrated at least modest improvement (10-30% reduction in delays and a rating of 3 out of 5 on the summary score); 11 teams were rated between 3-4, five teams were rated a 4, and eight teams were rated between 4-5.

Leape and colleagues summarized the fourth BTS project, "Reducing Adverse Drug Events and Medical Errors," in an article for the *Joint Commission Journal of Quality Improvement* in 2000. Forty teams met over 15 months to discuss and implement 12 change concepts (e.g., non-punitive reporting, standardized medication administration times). At the end of the study, 20% of hospitals had made one successful process change, 50% had made between two and four changes, 20% had made five or more changes, and 10% had dropped out of the project.[99] Although the major focus was on process changes, the article does present one example of the project's effect on clinical outcomes: a reduction in the number of adverse drug events in oncology and orthopedic units from 12 per month to 0-2 per month by the end of the collaborative period (no significance testing reported).

Two QICs in 1996, the fifth and sixth BTS projects, focused on administrative and clinical issues related to acute hospital care. "Reducing Costs and Improving Outcomes in Adult Cardiac Surgery" involved 41 teams working to meet eight specific targets for improvement

(e.g., "cath-to-CABG" time less than 48 hours, OR time for standard three-vessel CABG of no more than 180 minutes, OR costs less than $4,500 per case, 30-day mortality less than 2.0%).[100] To do this, the group implemented 13 changes to preoperative, intraoperative, and postoperative care along with the creation of cardiac surgery database for early- and long-term outcome tracking. Although there was no summary evaluation of teams, multiple examples of individual hospital success exist in the final report and one team went on to publish its own experience in a 1998 issue of *Quality Management in Healthcare*: a 30% reduction in length of stay, an increase in early postoperative extubation from 5% to 75%, a cost reduction of $19 per case, and an improvement in patients' pain, anxiety, service, and satisfaction scores.[101]

The other acute care QIC, "Reducing Costs and Improving Outcomes in Adult Intensive Care," involved 37 teams working to meet five clinical and administrative targets: a 30% reduction in the average time patients' spend on a ventilator, a 30% reduction in lab and x-ray costs, a 25% reduction in medication costs, a 25% reduction in length of stay, and a 40% reduction in ventilator-associated pneumonia.[102] Almost one-third of hospitals met or exceeded these goals over 12 months: 14 teams were rated between 3-4, six teams were rated a 4, three teams were rated between 4-5, and one team achieved a perfect 5 rating. Only one team published on its experience participating in the program, which included a reduction in ventilator use (exact data not reported), a commensurate increase in non-invasive ventilation (exact data not reported), a reduction in x-rays per patient day from 0.59 to 0.31, and a reduction in laboratory tests per patient day from 2.77 to 2.27, which accounted for almost $16,000 saved per year in this small, community hospital.[103]

Three of the final four QICs focused on chronic or end-of-life care. "Improving Care at the End of Life" (1997) involved 47 teams from hospitals, hospices, outpatient clinics, and home

care agencies. After 12 months, 42 teams had made "key changes" to their care processes (equivalent to a summary score of at least 3 out of 5), most commonly in pain/symptom management and advanced care planning.[104] Thirteen of these 42 teams made changes that were considered "substantial and measureable improvements" by the group (equivalent to a summary score of at least 4 out of 5), including a 60% reduction in patients reporting pain greater than 5 out of 10, a 100% increase in the number of patients with advanced care plans documented in the medical record, and a 100% increase in days of palliative or comfort care prior to death.[104]

In a second, related QIC, "Improving Care for Patients Approaching the End of Life with COPD and CHF" (1999), 34 teams from hospitals, hospices, and elder care organizations met to merge best practices from the medical care of patients with chronic obstructive pulmonary disorder or congestive heart failures with best practices from palliative care and symptom management. Twenty-nine teams made "key changes" over seven months while 16 made "substantial and measureable improvements," including a 25% reduction in exacerbations of heart or lung failure.[104] A matched cohort study of four sites participating in the collaborative and four organizationally similar controls (two hospitals, one health plan, and one physician group in both arms) found that QIC sites showed greater improvement in 11 of 21 CHF-specific process markers, particularly for education and counseling (+24% vs. -1%, $p<0.0001$).[87] Combining the process measures into a single score, QIC sites improved compliance by 17% over two years compared to only 1% for control sites ($p<0.0001$). Unfortunately, however, the study did not attempt to measure changes in clinical outcome.

Only one of the 31 teams participating in "Improving Care for People with Chronic Conditions" (1998) published data from its experience, although this specific team represented a group of 18 primary care clinics serving 170,000 adult patients.[105] The collaborative as a whole

focused on four components of chronic disease management (empowering patients in self-management, advancing the use of data, providing decision support to care teams, and redesigning care delivery systems), but each team was encouraged to apply these concepts to whichever patients they deemed most relevant. The team that published their results focused on diabetes care and found a 0.39-percentage-point reduction in hemoglobin A1c levels and a 4.3 mg/dL reduction in LDL levels on average for its 7,037 diabetic patients over 12 months ($p<0.001$ for both).[105]

The final BTS QIC with published data, "Improving Care for People Infected with HIV" (2000), was evaluated in a matched control study published by Landon and colleagues.[90] As was discussed earlier, this study, which showed improvements to HIV care processes but no additional benefit for QIC sites compared to controls, remains one of the few high-quality QIC evaluations to produce negative findings.

As a whole, I believe there are three major takeaways from the early BTS experience. First, the QIC method can be applied to--and can be successful among--a wide variety of topics. While this certainly does not represent definitive evidence in support of QICs, I do believe that it supports the idea that QICs can be thought of and evaluated more generally as a type of quality improvement tool rather than as a specific approach to a specific quality problem or topic. Second, hospitals that saw improvements did so for both clinical and administrative outcomes. Putting aside for one second whether or not these improvements can be attributed to QIC participation, this finding still differs from earlier experiences with CQI in which hospitals typically saw only administrative changes. It is possible, therefore, that aspects of the QIC approach (e.g., the conscious and deliberate inclusion of physicians on the improvement team) allow it to address clinical topics in ways that single-institution CQI programs could not. Finally,

success appears to be variable, both among QICs and among hospitals participating in the same QIC. Even the most successful BTS QICs demonstrated improvement among roughly two-thirds of its participants. Here, I believe two conclusions must be drawn. First, QICs--even QICs based on the same relatively structured model--are not perfectly homogenous and should not be evaluated as such ("if you've seen one QIC, then you've seen one QIC"[106]). Second, understanding the factors that drive this heterogeneity in success may allow us to design QICs that are more efficient and more likely to produce tangible improvements to quality.

QICs in surgery. Beyond my personal interest, there are several reasons to separately evaluate QICs that are designed to address surgical topics. First, the way in which surgeons collaborate, both with one another and with non-surgeons, is likely to be different from the way in which non-surgical clinicians collaborative, primarily due to the unique set of cultural and organizational norms associated with the OR.[107] For example, traditional hierarchies or the diversity of providers involved (e.g., surgeons, anesthesiologists, nurses) may complicate group dynamics, limit trust, and prevent the sharing of new ideas. Second, surgical patients are often more acutely or severely ill than the targets of other, non-surgical outpatient QICs (e.g., asthma, diabetes) or even non-surgical inpatient QICs (e.g., COPD exacerbation, labor and delivery). As multiple studies have suggested that QIC success may depend on the topic and target of the intervention,[40,62,108,109] the ability of surgical QICs to impact both processes of care and patient-level health outcomes may differ significantly from that of non-surgical collaboratives.

Therefore, to better understand the current landscape of QICs in surgery, I performed a systematic review of the literature to identify organizations that 1) met Schouten and colleagues' definition of a QIC, 2) focused on improving care in a single geographic region in the U.S. or Canada, and 3) had published on a topic related to general or vascular surgery (see Appendix 2.2

for search strategy). My initial search identified two major groups of QICs in surgery: 1) state-based QICs with either no connection or only a loose connection to ACS; and 2) state or provincial QICs that operate in partnership with ACS and use NSQIP as their platform for data collection and analysis. To generate a complete list, I supplemented my list of QICs identified by my literature search with a list of surgical QICs published in a narrative review by Maggard-Gibbons[110] and through personal communication with the person at NSQIP tasked with collaborative quality improvement efforts.[111] I then searched PubMed/MEDLINE again using both the names of individual QICs and their key personnel, who I separately identified either through NSQIP records or through QICs' official websites (see Appendix 2.3). My final list of QICs in surgery and a description of their organizational structure within the published literature is available in Table 2.2.

Overall, I found 19 examples of regional QICs in surgery: five state-based, independent QICs and 14 ACS-sponsored, NSQIP QICs. I identified a total of seven QICs meeting my inclusion criteria via literature search, including all five independent QICs. Of note, I found several articles from QICs in other surgical sub-specialties (e.g., cardiac surgery, pediatric surgery, urology), four international QICs (two in the Netherlands,[112,113] one in the UK, and one in Australia[114]), and two non-regional U.S. QICs, all of which I excluded based on my pre-set criteria. Both nationwide U.S. QICs focused on SSI and have been cited repeatedly in the literature. Dellinger and colleagues[115] found improvement to both anti-microbial processes and overall infection rates, but did not include a control group. Kritchevsky and colleagues[116] also found higher compliance with process measures, but no difference between hospitals randomized to participate in a QIC and those randomized to receive feedback only. This final study was also included in both Schouten and colleagues' and Nadeem and colleagues' systematic reviews.

I also identified 29 NSQIP QICs from either Maggard-Gibbons (2014) or through direct communication with NSQIP; both NSQIP QICs that I identified via my literature search were also listed in these supplemental sources. To be included in the official NSQIP registry, QICs must meet four criteria: 1) two or more hospitals, 2) an identified leader or leadership team, 3) a written data-sharing agreement, and 4) a written or verbal mission statement or charter.[111] Of the 29 NSQIP QICs identified by supplemental sources, I excluded 15 either because they did not meet Schouten and colleagues' definition of a QIC (13) or because they did not take place within a single geographic region (2; see Appendix 2.4).

Two of the five independent QICs have published extensively on their activities, although only two studies from the same QIC have attempted to evaluate the collaborative model using a quasi-experimental design.[117,118] None of the NSQIP QICs have formally evaluated their collaborative using experimental or quasi-experimental designs and only three present data on changes in either clinical processes or patient outcomes over time.

The first independent QIC in surgery was actually more of an isolated implementation project than a sustained collaborative intervention. In 1993, the Health Care Financing Administration began is Health Care Quality Improvement Program (HCQIP), which included state-wide evaluations of certain care processes. As a part of HCQIP implementation, Wisconsin performed its own internal review of acute care hospitals and found that the vast majority did not have a policy in place for meeting national standards on post-operative pain relief. In response, the state, through its peer review organization MetaStar, invited all acute care hospitals in the state to participate in a collaborative effort designed to disseminate evidence-based guidelines. Seventeen hospitals agreed to participate and submitted baseline data. As a part of the collaborative, these hospitals attended seminars with content experts in pain management and

Table 2.2: Survey of QICs in Surgery and their Organizational Structures

Name, date founded	Location, description of participant hospitals	Structure, funding	Topic, disease, or condition	Features/Activities	Published examples of change over time
MetaStar HCQIP Collaborative*[119] 1995	• Entire state of Wisconsin • Voluntary (17 hospitals)	• Administered by MetaStar (a peer review organization in Wisconsin)	• Post-operative pain management	• Independent data collection and analysis • Development and tracking of novel process measures (e.g., route of pain medication administration, use of non-pharmacological pain control) • Multi-disciplinary expert panel with seminars on evidence-based guidelines • Feedback of statewide baseline data • Training in quality improvement methodology	• Reduced use of intramuscular pain medicine and meperidine[119] • Reduction in self-reported pain scales[119] • Increased use of scheduled dosing and increased frequency of pain assessment[119]
Michigan Surgical Quality Collaborative (MSQC)[120-122] 2005	• Entire state of Michigan • Voluntary, but incentivized by largest private insurer (15 hospitals at launch, currently 66 hospitals) • Mostly small and rural community hospitals	• Administered by the University of Michigan • Lead by executive committee of 5 surgeons (3 from community centers and 2 from academic centers) • Previously associated with the ACS • Funded by BCBS of Michigan and Blue Care Network	• All general or vascular surgical procedures Specific projects include: • Colectomy -Prophylactic antibiotics -Bowel preparation -Anastomotic leak -Readmissions • Emergency surgery • Surgery in the elderly • Post-operative hypoglycemia • Failure to rescue	• Independent data collection and analysis • Development and tracking of novel process measures (e.g., rates and types of pre-operative testing for patients with a history of myocardial infarction[121]) • Outcome benchmarking via quarterly reports (de-identified data) • Communication between providers (quarterly meetings, electronic message board, paper newsletter, YouTube site) • Site visits • Initiatives for specific conditions or operations	• Reduced rate of 30-day, risk-adjusted morbidity after surgery, faster rate of improvement than among other NSQIP hospitals[118] • Increased compliance with perioperative antibiotic prophylaxis in colectomy,[128,129] but not in post-operative SSI rates[130] • Reduced rate of complications following vascular surgery[131] • Reduced rate of major complications (e.g., myocardial infarction,

				(colectomy,[123] lower-extremity revascularization) • Educational and/or intervention activities (best practices for reducing SSI after colorectal surgery[121,124-127]) • Financial incentives for participating ("pay-for-participation")	stroke) following pancreatic surgery[132]
Surgical Care and Outcomes Assessment Program (SCOAP)[122,133,134] 2006	• Entire state of Washington • Voluntary with the goal of universal participation (17 hospitals at launch, currently 45 hospitals[135])	• Administered by the Foundation for Healthcare Quality • Lead by a committee of surgeons and quality improvement leaders • Associated with the Washington State Chapter of the ACS • Funded by "multiple stakeholders including purchasers, payers, and hospitals", including the Washington State Health Care Authority	• Initially: colon and rectal surgery, bariatric surgery, and appendectomy • Now includes: vascular surgery interventional vascular procedures, pediatric surgery, spine surgery, gynecologic surgery, urology, high-risk cancers (esophagectomy, pancreatectomy, hepatectomy), and outpatient surgery (hernia, cholecystectomy, breast) Specific projects include: • Negative appendectomy • SSI • Perioperative glycemic control • Appropriateness of elective colon	• Independent data collection and analysis • Development and tracking of novel process measures (e.g., accuracy of diagnostic imagining for appendicitis, appropriate use of neoadjuvant therapy for rectal cancer) • Outcome benchmarking via quarterly reports (de-identified data) • Communication between providers (annual meeting, e-newsletters) • Educational and/or intervention activities (standardized orders and OR checklist)	• Decreased rate of negative appendectomy, particularly for high-risk populations[136-138] • Increased rate of intra-operative leak testing and decreased rate of re-intervention or death in elective colorectal surgery[139,140] • Increased continuity of perioperative β-blocker administration[141] • Increased use of laparoscopy in colorectal surgery[142] • Decreased cost per case for appendectomy, colorectal, and bariatric operations[134] • Increased rate of appropriate elective colorectal surgery in diverticulitis[143] • Increased rate of blood glucose testing of diabetic patients undergoing elective colorectal surgery[134]

			resections for diverticulitis		• Increased use of chemoprophylaxis for VTE in colorectal surgery[144,145]
Michigan Breast Oncology Quality Initiative (MiBOQI)[146] 2006	• Entire state of Michigan • Voluntary (18 participants out of 42 qualifying centers) • Sites must have sufficient volume to participate (100+ breast cancer cases per year)	• Administered by the University of Michigan • NCCN Breast Cancer Outcomes Project Database • Funded by BCBS of Michigan	• Breast cancer	• Independent data collection and analysis • Outcome benchmarking via quarterly reports (de-identified data) • In-person meetings 3 times per year	• Decrease in the percentage of patients undergoing surgical as opposed to core-needle biopsy for suspected breast cancer[146]
Upstate New York Surgical Quality Initiative (UNYSQI)[147,148] 2007	• Upstate New York • Voluntary (8 hospitals at launch, currently 12 hospitals)	• Associated with the ACS • Funded by Excellus BCBS	• All NSQIP-recorded operations • Anastomotic leak after colectomy	• Independent data collection and analysis • Development of novel quality indictors (anastomotic leak after colectomy[147]) • Outcome benchmarking • Initiatives for specific conditions (colectomy)	• None to date
Tennessee Surgical Quality Collaborative (TSQC)[110,149-151] 2008 (first data collected in Jan 2009)	• Entire state of Tennessee • Voluntary (10 hospitals at launch, currently 21 hospitals)	• Administered by the Tennessee Hospital Association through the Tennessee Center for Patient Safety • Leadership committee includes surgeons, hospital CEOs, and a member of the Tennessee Hospital Association • Associated with the Tennessee Chapter of the ACS	• All NSQIP-recorded operations Specific topics: -Surgeon-specific reporting -Operative time	• Independent data analysis • Outcome benchmarking • Access to best practice guidelines and case studies through NSQIP • Initiatives for specific conditions (reducing UTI rates, improving colorectal surgery outcomes)	• Decreased rates of acute renal failure, graft failure, ventilator use, superficial SSI, and wound disruption, but increased rates of DVT, pneumonia, and UTI[150] • Cost savings of $2 million per 10,000 cases in the first 2 years[150]

		• Funded by the Tennessee BCBS Foundation			
Florida Surgical Care Initiative (FSCI)[152-156] 2008 (first data collected in Mar 2011)	• Entire state of Florida • Voluntary, but previously incentivized by a large private insurer (54 hospitals at launch, rose to 102 hospitals at peak, currently 24 hospitals)	• Administered by the Florida Hospital Association • Associated with the ACS • Previously funded in part by BCBS of Florida • Endorsed by the Florida Chapter of the ACS, the National Patient Safety Foundation, the Florida Health Care Coalition, Mayo Clinic, and the Institute for Healthcare Improvement	• NSQIP-recorded general and vascular surgery operations	• Outcome benchmarking via semi-annual reports on a limited number of measures (SSI, UTI, colorectal complications, elderly surgery complications) • Communication between providers (meetings, monthly conference calls, educational webinars) • Access to best practice guidelines and case studies through NSQIP • Voluntary participation in the Johns Hopkins Surgical Unit-based Safety Program	• Decreased rates of UTI, SSI, and complications in colorectal and elderly surgery patients over 15 months[156] • Cost savings of $6.7 million over 15 months[156]
Pennsylvania National Surgical Quality Improvement Program (PA-NSQIP)[157] 2010	• Entire state of Pennsylvania • Voluntary (9 hospitals)	• Administered by the Pennsylvania Patient Safety Authority • Steering committee of surgeons, nurse reviewers, and Authority staff (patient safety liaison and infection prevention specialist)	• General, vascular, and colorectal surgery • SSI after bariatric and colorectal surgery	• Communication between providers (workspace on website) • Site visits by physicians and other clinical staff, process benchmarking via interviews and collaborative discussion • Educational and/or intervention activities (targeted education on SSI prevention, development and validation of a unique SSI prevention assessment tool)	• None to date
Western Pennsylvania SSI Collaboration/ Pennsylvania	• Entire state of Pennsylvania • Voluntary (40 hospitals)	• Administered by the Hospital and Healthsystem	• All elective surgical procedures tracked by the National	• Communication between providers (monthly conference calls and webinars)	• Reduction in the overall standardized infection ratio with the largest effect in colon, cesarean

Hospital Engagement Network (PA-HEN)[158] 2010		Association of Pennsylvania • Associated with the Pennsylvania Patient Safety Authority	Healthcare Safety Network	• Educational and/or intervention activities (screening and decolonization protocol for patients admitted for elective surgery) • Involvement of patients in the quality improvement process	section, hip replacement, knee replacement, and laminectomy[158]
British Columbia National Surgical Quality Improvement Program (BC-NSQIP)[159-161] 2011	• Initially hospitals in British Columbia, now expanded to the entire country of Canada (CAN-NSQIP) • Voluntary (25 hospitals) • 90% of surgical procedures in British Columbia	• Coordinated by the British Columbia Surgical Quality Action Network • Associated with the British Columbia Patient Safety and Quality Council • Funded by the British Columbia Health Services Purchasing Organization	• All NSQIP-recorded operations	• Independent data analysis • Outcome benchmarking via quarterly reports (two from ACS, two from collaborative) • Initiatives for specific conditions (UTI, SSI, pneumonia) • Opportunity to participate in the British Columbia Surgical Quality Action Network ("a forum for health care providers to discuss best practice, share local innovations, and connect"[162])	• Decreased rate of UTI for the entire province of British Columbia over 2 years[160] • Individual hospital examples of a reduction in rates of UTI, SSI, pneumonia, and overall morbidity after surgery[160,161]
Connecticut Surgical Quality Collaborative (CtSQC)[163-165] 2011	• Entire state of Connecticut • Voluntary (4 surgeons at launch, now 20 hospitals)	• Sponsored by Coverys (a medical malpractice insurance provider) • Partnered with Connecticut Chapter of the ACS, Connecticut Hospital Association, Connecticut State Medical Society,	• Colorectal surgery (participation in NSQIP is not required) Specific topics: • ERAS • SSI • Readmission • Transfusions	• Communication between providers (quarterly meetings) • Educational and/or intervention activities (presentations on quality and best practices) • Multidisciplinary team (surgeons, anesthesia, nursing [OR, PACU, bedside], home care, nutrition, respiratory therapy, PT/OT, informatics) • In-person meetings • Process benchmarking (high-performing centers train other centers) • Outcome benchmarking • Bundles	• None to date

Georgia Surgical Quality Collaborative (GaSQC)[166] 2012	• Entire state of Georgia • Voluntary (8 hospitals at launch, now 18 hospitals)	• Associated with the Georgia Society of the American College of Surgeons	• All NSQIP-recorded operations • SSI	• Outcome benchmarking • CUSP program • Development of best practice guidelines in four phases (at home, pre-op, intra-op, and recovery) • Lean • Multidisciplinary team • Integration with HER • PDCA • Clinical and administrative	• None to date
Illinois Surgical Quality Improvement Collaborative (ISQIC)[167,168] 2015	• Entire state of Illinois • Voluntary (54 hospitals at launch)	• Administered by the Surgical Outcomes and Quality Improvement Center at Northwestern University • Associated with the Illinois and Metropolitan Chicago Chapters of the ACS, ACS-NSQIP, and BCBS of Illinois • Advisory committee with representatives from local ACS Chapters, ACS NSQIP, coordinating center, BCBS of Illinois, hospitals	• All NSQIP-recorded operations Specific topics: • VTE prophylaxis	• Outcome benchmarking (hospital and surgeon level) • Process benchmarking (site visits) • Communication between providers (3 annual meetings, monthly webinars) • Training in quality improvement methodology (formal curriculum, process improvement coach) • Financial incentives to improve outcomes • Formal evaluation of interventions	• None to date

*Descriptive title; no official name listed

Five regional QICs listed in Maggard-Gibbons (2014)[110] or in ACS NSQIP documentation that could not be associated with either a publication or an official website include: Northern California Surgical Quality Collaborative, Nebraska Collaborative, Oregon NSQIP Consortium, Virginal Surgical Quality Collaborative, Ontario Collaborative

ACS, American College of Surgeons; BCBS, Blue Cross Blue Shield; CEO, Chief Executive Office; DVT, deep venous thrombosis; HCQIP, Health Care Quality Improvement Program; NCCN, National Comprehensive Cancer Network; NSQIP, National Surgical Quality Improvement Program; SSI, surgical site infection; UTI, urinary tract infection; VTE, venous thromboembolism

then returned to their institutions to implement their own quality improvement interventions. Six months after the final seminars, hospitals were again asked to submitted data on their peri-operative pain management practices. Although hospitals increased their compliance, on average, with all six process measures, there was wide variation in success among the sites: only two hospitals were able to decrease their use of intramuscular pain medications compared to seven hospitals that decreased their average patient-reported pain scale scores, the largest number for any quality indicator.[119] Interestingly, two hospitals actually had a statistically significant reduction in the frequency of pain assessment and four hospitals had a statistically significant reduction in the use of non-pharmacological pain control, both of which were in the *opposite* direction of the intended effect.

The second independent QIC, the Michigan Surgical Quality Consortium (MSQC), actually began in 2005 as a partnership between ACS, Blue Cross Blue Shield of Michigan (BCBSM), and 16 regional hospitals, but separated from ACS in order to have more control of their collaborative process and to collect additional data outside of the NSQIP system.[121] Currently, MSQC is administered out of the University of Michigan with BCBSM serving as a third-party sponsor. In this capacity, BCBSM covers all of the administrative costs of the collaborative in exchange for hospitals' commitment to improve quality and their creation of semi-annual, de-identified reports on system-wide clinical outcomes--a strategy the group has dubbed "pay for participation."[169] An analysis by Englesbe and colleagues at the outset of MSQC suggested that a reduction in the statewide complication rate of 1.8% per year over the first three years would make the program cost neutral to BCBSM while a reduction of 3% per year--the same reduction that was demonstrated by original NSQIP studies--would actually net BCBSM $2.5 million over the same period.[170]

Two studies have used quasi-experimental approaches to evaluate the program. Campbell and colleagues (2010) compared mortality and complication rates between 16 MSQC hospitals and 126 Michigan hospitals that did not participate in MSQC, but did participate in NSQIP. Over 16 months, MSQC hospitals reduced their complication rates from 10.7% to 9.7% (p=0.002) despite there being no changes in complications rates for the comparison group (12.4% vs. 12.5%, p=0.49). These trends corresponded to a 10.2% risk-adjusted reduction in the odds of having a complication in the MSQC hospitals compared to control hospitals (p=0.004); there was no difference in risk-adjusted mortality rates over time in either group.[117] In a similar study, Share and colleagues found that risk-adjusted complication rates in MSQC hospitals fell from 13.1% in 2005 to 10.5% in 2009 (p<0.001) while rates for NSQIP hospitals outside of Michigan showed no change over the same period.[118] Overall, complication rates in MSQC hospitals were lower (10.5% vs. 11.5%, p<0.001) and improved at a faster rate than NSQIP comparison hospitals (p<0.001).

Several other studies have tracked changes in process measures or patient outcomes over time for MSQC hospitals, but did not use a control group for comparison. Campbell and colleagues (2009) document an increase in compliance with 10 novel process measures for SSI over time, but do not perform statistical testing.[121] This same study comments that SSI rates had not improved over time, but does not present data to support this claim. Henke and colleagues found a lower risk-adjusted morbidity rate (15.8% vs. 13.8%, p=0.02), a 15% reduction in length of stay (p<0.001), and a 11% reduction in operative time (p<0.001) for open vascular operations after MSQC was established.[131] Smith and colleagues found that morbidity and mortality rates were "relatively unchanged" for MSQC hospitals between 2005 and 2010, but did not provide statistical testing.[171] Healy and colleagues found differences in risk-adjusted complication rates,

mortality rates, and failure-to-rescue rates between low-caseload and high-caseload hospitals for pancreatic resection prior to MSQC; these differences disappeared, however, within three years after collaborative formation.[132] The overall variation in risk-adjusted complication rates decreased 44% over the same time period. Finally, Reames and colleagues compared complication rates among MSQC hospitals implementing a separate quality initiative, the Keystone Surgery checklist program, to MSQC hospitals not implementing the program over four years. The authors found no difference in risk-adjusted complication rates; however, they do not report changes in checklist adherence to determine if compliance rates changes over time.[130] It is unclear whether high pre-intervention compliance with recommended practices as a part of MSQC participation played a role in the study's negative findings; a poor choice of process measures or incomplete implementation would have similarly resulted in no change to clinical outcome.

Three factors, in particular, may have contributed to the apparent success of the MSQC program. First, two previous multi-institutional efforts involving the same state and many of the same hospitals--the Blue Cross Blue Shield Cardiovascular Consortium of Michigan[172] and the Keystone ICU project[173]--had been successful at changing practices and improving clinical outcomes. These real, clinical success stories may have convinced clinicians, the primary holdout during CQI-based efforts, of the value of quality improvement. Second, MSQC's "pay for participation" arrangement with BCBCM ensured that membership was cost neutral at the hospital level, which gave administrators and leadership fewer reasons not participate.[170] Finally, the agreement among hospitals and BCBSM that data would remain de-identified and could not be used for public reporting or hospital marketing allowed skeptical providers to participate without fear of public embarrassment or financial consequences.[120]

The third independent QIC, the Surgical Care and Outcomes Assessment Program (SCOAP), was founded in 2006 and has since expanded to include 40 hospitals across Washington State. Like MSQC, SCOAP is supported financially by payers (including the state's Medicaid authority) and uses an external third-party, the Foundation for Health Care Quality, to guarantee a safe environment for exchange by maintaining the anonymity of hospital outcomes. SCOAP does not allow public reporting of hospital-level data or the use of program data for marketing purposes.

Also similar to MSQC, two previous projects paved the wave for the formation of SCOAP. First, the major stakeholders in SCOAP had worked together previously to create a statewide clinical registry for percutaneous and surgical cardiac interventions, the Clinical Outcomes Assessment Program (COAP). More similar to NSQIP than a QIC, COAP rigorously collects clinical data on patients undergoing CABG surgery, produces reports, and feeds back information, but it relies on individual hospitals to generate its own quality improvement activities.[133] Unlike NSQIP, participation in COAP (and SCOAP) is paid for, in part, by external organizations, allowing for greater participation from smaller, regional hospitals and a more comprehensive view of surgical care in the state.[134] In a direct precursor to SCOAP, researchers at the University of Washington also performed a surgical needs assessment for the state, documenting widespread variation in outcomes and calculating a potential $30 million per year cost savings that could be generated by bringing above-average complication rates and hospital days down to the state average.[133]

No formal evaluation of SCOAP using either an experimental or quasi-experimental design has been published to date. There are, however, several examples of changes in processes or outcomes over time in areas targeted by the program. Cuschieri and colleagues found that

rates of negative appendectomy--removing the appendix for suspected appendicitis only to find that it is pathologically normal--decreased from 7% to 5% within one year of SCOAP encouraging its hospitals to increase their use of pre-operative imaging.[136] (Rates of preoperative imaging for suspected appendicitis also rose in parallel, based on two studies by Kwon and colleagues [2012 A][134] and Drake and colleagues.[137]) Kwon and colleagues (2011) found an increase use of pharmacologic prophylaxis for venous thromboembolism (35.8% vs. 70.4%, p=0.001) and an associated reduction in 90-day venous thromboembolism rates (4.3% vs. 1.7%, p=0.001) over the first three years of the program (2005-2008).[144] These findings were updated by Nelson and colleagues to include data through 2011; while the use of pharmacologic prophylaxis continued to rise (86.4%, p<0.001), there was no change in the 90-day rate of venous thromboembolism during the most recent period (3.0%, p=0.09).[145] Kwon and colleagues (2012 A) found increased rates of anastomotic leak testing (also shown by Kwon and colleagues [2012 B] with associated reductions in re-intervention for leak[139]), increased rates of blood glucose checks for diabetic patients, and decreased variability in complication rates following elective colorectal resections (no statistical testing performed).[134] The same study also found a decreased rate of adverse events (not defined) over time and a roughly $6,000 lower cost per case for appendectomy, colorectal, and bariatric operations for SCOAP hospitals compared to non-SCOAP hospitals in 2009 (comparison group not identified). Finally, Kwon and colleagues (2012 C) found higher rates of perioperative β-blocker continuation (57.2% vs. 71.3%, p<0.001) after one year of participation in SCOAP.[141]

In addition to changes in clinical parameters over time, two papers document organizational changes encouraged by SCOAP. Kwon and colleagues (2012 D) found an increased use of laparoscopy in elective colorectal resection (23.3% vs. 41.6% of cases,

p<0.001), another SCOAP target, over the first five years of the program.[142] Simianu and colleagues found that the proportion of patients undergoing an elective colon resection for diverticulitis without an evidence-based indication for surgery decreased from 38.4% to 26.4% over four years (p<0.001).[143] SCOAP had previously generated evidence-based guidelines for operating in diverticulitis and began feeding back data to its member hospitals on their rates of inappropriate colon resection starting in 2011.

Another Michigan-based QIC, the Michigan Breast Oncology Quality Initiative (MiBOQI), was also formed in 2006 as a statewide, BCBSM-funded partnership for improving the care of women with breast cancer. Like MSQC, MiBOQI is coordinated out of the University of Michigan and is a voluntary, but requires centers to meet certain volume and data quality standard to participate. Unlike MSQC, which initially relied on NSQIP and then began to collect its own data, MiBOQI uses clinical data submitted to the National Comprehensive Cancer Network Breast Cancer Outcome Project Database. In the only published article from the group, Breslin and colleagues report trends in surgical as opposed to core needle biopsy rates for suspected breast cancer, an initial focus on the groups collaborative discussions. The authors demonstrate a significant reduction in surgical biopsy rates from 20% of patients to 15% over a three-year period.[146] As is the case with other QICs, there was substantial variation among sites, with 8% of cases receiving a surgical biopsy at one site compared with 37% at another.

The one other state-based QIC, the Pennsylvania Hospital Engagement Network (PA-HEN), was formed in 2010 as a partnership between the Hospital and Healthsystem Association of Pennsylvania, the Pennsylvania Patient Safety Authority (an independent state agency), and two independent contractors. Like MSQC and SCOAP, PA-HEN directly built on an earlier collaborative effort, the Western Pennsylvania SSI Collaboration, which involved group

discussion and the sharing of best practices among 19 hospitals. PA-HEN expanded this collaboration to include 40 hospitals, spread the effort across the entire state, and focused on improving patient engagement. The final intervention consisted of patient education about SSI during their initial office visit, pre-operative decolonization using chlorhexidine baths and muciprocin nasal swabbing, and provider education to improve intra-operative infection control. The standardized infection ratio for elective cases among these hospitals consistently decreased over the intervention period, from 1.27 in 2010 to 1.17 in 2011, 0.797 in 2012, and 0.735 in 2013.[158] Reductions were seen for the majority of elective cases, except for CABG, which had already instituted a protocol for pre-operative decolonization. No data were provided about differences in infection rates across hospitals, patient compliance with pre-operative decolonization, or patient or provider perceptions of the program.

Only three NSQIP QICs have published longitudinal data to date: two in peer-reviewed journals and one in an annual report. The Tennessee Surgical Quality Collaborative (TSQC) was formed in 2008 as a partnership between the Tennessee Hospital Association, Blue Cross Blue Shield of Tennessee, and 10 hospitals. Like MSQC, Blue Cross Blue Shield provides funding to run the collaborative, although, in TSQC, the Tennessee Hospital Association acts as the third-party coordinating center and is the only organization with access to de-identified data from all hospitals. An analysis by Guillamondegui and colleagues comparing post-operative complication rates before and after the formation of TSQC found mixed results. Risk-adjusted rates of acute renal failure (75.3 vs. 56.4 per 10,000 procedures, $p=0.023$), graft/flap failure (45.8 vs. 18.1, $p<0.0001$), prolonged ventilation (over 48 hours; 293.6 vs. 250.3, $p=0.012$), SSI (357.6 vs. 289.9, $p<0.001$), and wound disruption (90.8 vs. 59.7, $p=0.011$) all decreased in the year following QIC formation; however, rates of deep vein thrombosis/thrombophlebitis (66.2 vs.

89.3, p=0.013), pneumonia (224.6 vs. 276.5, p=0.002), and urinary tract infections (UTIs; 164.7 vs. 233.5, p<0.0001) all increased during the same period.[150] Despite the mixed results, a financial analysis based these changes in complication rates suggested an almost $4.5 million savings to TSQC hospitals over the same time period.

A second NSQIP QIC reporting data, the Florida Surgical Care Initiative (FSCI), formed in 2011 as a partnership between the Florida Hospital Association, which administers the collaborative, Blue Cross Blue Shield of Florida, which provided the initial funding, and 54 hospitals. As only six of the 54 hospitals were NSQIP members at the outset of the collaborative, FSCI decided to partner with ACS and provide a limited version of program (termed "NSQIP-lite") for a select group of general and vascular surgery patients. Tepas and colleagues compared FSCI hospitals' performance on four outcome measures over six months. Overall, the hospitals had a 14.5% reduction in complications, which accounted for almost $7 million and 89 lives saved.[156] The percentage of hospitals meeting their quality goal--defined as an observed-to-expected ratio less than 1.0--increased for catheter-associated UTIs, adverse events among patients undergoing colorectal surgery, and adverse events among patients over 65 years old; however, the percentage of hospitals meeting their quality goal actually decreased for SSI. No statistical testing was performed and almost no information was provided about the variability of results among hospitals, although the authors note that "only 3 of 54 hospitals demonstrated improvement across all 4 measures."[156(p602)]

The final NSQIP QIC reporting data, the British Columbia National Surgical Quality Improvement Program (BC-NSQIP), is actually part of a larger collaborative effort in surgery, the Surgical Quality Action Network (SQAN), which is administered by the British Columbia Patient Safety and Quality Council and funded by the British Columbia Health Services

Purchasing Organization, an agency of the provincial government. In addition to BC-NSQIP, SQAN manages two working groups: one on teamwork and communication and a second on clinical care management, which has focused on implementing surgical safety checklists and promoting best practices for SSI.[162] Ten of the 24 hospitals participating in BC-NSQIP were identified as needing improvement based on their initial UTI rates and the entire group immediately began local and coordinated quality improvement efforts. After one year, these 10 hospitals dropped their UTI rates from 3% to 2.25%, but were unable to match the overall BC-NSQIP average of 1.3% (no statistical testing reported, although the report states that there were "not enough data points to say that we have made a significant improvement"[160(p6)]).

Brief summary of the literature and its limitations

Given the amount of literature presented, I believe a brief summary of *whether* QICs work in quality improvement is warranted before presenting evidence on *how* they may do so. As with many areas of quality improvement,[174] the evidence for QICs is, in general, inconsistent, limited, and of moderate-to-poor quality. As a movement, CQI--QICs clear substantive and temporal precursor--appears to have influenced the way we think and talk about quality and quality improvement, despite little-to-no empiric evidence that CQI-based interventions are an effective, much less efficient, method for improving quality of care. Several systematic reviews in surgery do provide some support for CQI-based interventions, but is it unclear whether the specifics of surgical disease (e.g., the structure and culture of the OR, the acuity and severity of disease) or the hospitals involved (i.e., hospitals that would have improved anyway) played a role in these findings.

Perhaps because of the large-scale reorganization the CQI approach often requires, success under CQI appears to be particularly dependent on having a local context that supports the process, which, again, given the large-scale change, is particularly difficult to find. Organizations that failed to align quality improvement work with their other strategic goals found themselves unable to generate and sustain support from clinicians. Without this support, CQI came to look more like quality assurance than quality improvement and was quickly rejected in many settings.

Building on this frustration, QICs were designed to provide better contextual environments for CQI-based interventions by directly connecting clinicians and senior leadership, by developing reasonable expectations for improvement, by establishing an abbreviated timeline, and by leveraging support (and pressure) from other institutions to help maintain tension for change. In at least some instances, this strategy appears to have worked. The NNECVSG, NIC/Q, multiple BTS projects, and at least eight regional QICs in surgery have published evidence of improvement over time, although the majority of these articles rely on uncontrolled, pre-post study designs. More formal evaluations using experimental or quasi-experimental designs are generally less supportive of QICs as a relative improvement over single-institution quality improvement strategies, although several examples of QIC hospitals outperforming non-QIC hospitals do exist. As the concept of QICs retains its face validity and no study has definitively reported negative effects from QICs, the question of whether an organization should participate in a collaborative appears to be one of comparative effectiveness rather than risk/benefit. Do QICs provide enough "bang for the buck" in terms of lower complication rates, fewer readmissions, and fewer patient deaths, especially given their typically higher price tag in financial, organizational, and temporal resources?

Unfortunately, this and several other questions about the QIC methodology remain mostly unanswered by the current literature. Most glaringly, no one--not even the senior leadership at IHI[65]--has developed a convincing model for explaining how QICs work, why some work better than others, and why some organizations within a QIC meet their goals while others fail to improve.[65] Nor has any study suggested ways in which QICs, acting as local laboratories for quality improvement, might provide a framework for testing new quality improvement strategies or for adapting current practices to a specific local or regional context. Building a framework for thinking about and discussing these concepts--even if this discussion remains mostly conceptual--is the primary goal of my original work.

Chapter 3 Preface to My Original Research

Despite significant efforts to understand and enhance the QIC approach since the late 1990s, there remain substantial gaps in the literature, as I demonstrated in my previous chapter. What I believe the literature *does* offer is two somewhat-disconnected pieces of the larger narrative on QICs. First, there are dozens of published descriptions of QICs and several seminal reviews of the QIC approach, which, taken together, generate a relatively good picture of *what QICs do*: teams from multiple hospitals--often with the help of clinical and quality experts--collaborate, both in person and through remote networks, in the hope of identifying, sharing, and implementing clinical best practices. Second, there is a smaller series of studies that attempts to connect participation in a QIC to improvements in clinical quality. Putting aside the methodological challenges, this second group of studies sheds some light on *whether QICs work*: probably for particular groups of hospitals facing particular quality problems, but not for others. What is largely missing from the literature is any attempt to connect these two concepts, specifically an explanation of *how and why QICs work*. What is the theory of change that connects participating in collaborative group meetings to changes in hospital processes and patient-level health outcomes?

As I explain in the following chapter, this important question remains unanswered largely because of the way previous research has approached QICs. Too much time has been spent asking whether QICs work--itself a misleadingly binary question--instead of trying to understand how participation in a QIC might affect how hospitals carry out quality improvement activities and how these activities might, in turn, affect patients' health. Put another way, previous work has jumped to measuring the end-impact of a particular intervention (i.e., the QIC) without establishing a framework for understanding how the intervention might produce its expected

outcomes. Without such a framework, it is more difficult to attribute observed improvements in quality to the intervention itself and nearly impossible to distinguish failures of concept from failures of implementation when no obvious improvements occur.

This final point--understanding the causes and sources of QIC failure--is particularly important for guiding hospitals toward the most efficient and effective quality improvement strategies. If the heterogeneity in success we have seen among previous QICs is a failure of concept (i.e., the QIC method itself adds no value), then the ultimate solution may be to scrap the entire approach and push hospitals toward other quality improvement tools, such as clinical registries or single-institution CQI. If, however, this heterogeneity reflects a failure of implementation (i.e., the method adds value, but only if applied correctly), then the solution may be to more selectively apply the QIC approach, to focus on aspects of the approach that are particularly important to success, and to modify the approach to fit the specific set of hospitals and quality problems within each individual QIC. My hypothesis is that this latter type of failure--incomplete or imperfect implementation--is the predominant force behind QICs' variable success to date, just as it was with CQI several decades before.

To explore this idea further, I developed a novel conceptual model in Chapter 4 that attempts to connect a hospital's decision to participate in a QIC to changes in the way it performs quality improvement and, eventually, changes in its patients' health outcomes. After reviewing the literature and interviewing several key contributors to the QIC methodology, I decided that I could best answer the larger question of *how and why QICs work* by breaking it into three smaller but interconnected sub-questions. First, how does participation in a QIC affect the way an organization envisions and carries out its quality improvement activities? Second, which QIC components and processes are important to ensuring that hospitals will be positively affected by

their participation (i.e., what makes the collaborative process itself work well)? And, finally, how might we assess and categorize QICs that do not succeed in order to better understand and correct aspects of the collaborative process?

By answering the first sub-question, I hoped to advance our ability to evaluate and adapt QICs as we would other quality improvement interventions, namely by developing a logic model and following the decision to participate in a QIC through its intervening effects on hospitals until the pathway eventually reaches processes that affect patient health. My hope is that this model will eventually provide a new method for evaluating QIC implementation and for identifying common barriers to translating participation into better patient health.

By answering the second sub-question, I hoped to learn how to design more effective and efficient QICs. For example, much has been written recently on the length of the collaborative process (i.e., shorter versus longer period) and the medium of collaboration (e.g., in-person versus virtual QIC meeting). If certain changes can be made to preserve the benefits of collaboration while decreasing the cost of the program, then my research may help to extend the reach of QICs and to re-pique the interest of hospitals that chose not to participate previously because of the expected costs.

Finally, by answering the third sub-question, I hoped to be able to more thoroughly explain the heterogeneity in success that has come to mark the literature on QICs. My hypothesis here, and a main finding of Chapter 4, is that QICs may succeed and fail in at least two separate ways. First, the collaborative process itself may fail (e.g., hospitals refuse to meet or refuse to share their ideas). In this case, no hospital gains additional benefit and organizers must focus on modifying QIC-level factors when devising a solution (e.g., the number of hospitals, the type of hospitals, the organizational structure of the meetings). Second, the collaborative process may

function, but the benefits it produces (e.g., technical support, quality training) may only be useful to a subset of participating hospitals. In this case, organizers must either modify hospital-level factors so that more sites will benefit (e.g., changing how interventions are implemented at member hospitals or working to influence local culture) or restrict QIC participation to hospitals that they think will benefit most. While my current work stops short of developing a tool or checklist for ensuring that QICs succeed, my hope is that what I have been able to discover so far will set the stage for future research in this area.

After exploring *how and why QICs work* in Chapter 4, I chose to transition slightly in Chapters 5 and 6 in order to use data from our own QIC to answer two important corollary questions: how do QICs fit into the larger landscape of quality improvement work and what role might they play in advancing quality improvement more generally? Although each chapter uses data from LACTC to explore a unique contribution of QICs, both chapters examine the relationship between QICs and external standards--Chapter 5 deals with accuracy of TQIP's method of risk adjustment for patients with TBI while Chapter 6 deals with the appropriateness of implementing national guidelines for ICP monitoring among TBI patients in LA County.

Chapter 5 explores the importance of outcome benchmarking, specifically what role QICs might play in collecting new data and developing new models for risk adjustment. Unlike national programs, QICs are typically smaller, more flexible, and better equipped for experimenting with data collection practices. I argue that, by collecting additional clinical variables, QICs might provide two unique benefits--not only to their member hospitals, but to all hospitals interested in quality improvement. First, since benchmarking is based on identifying and learning from high-performing centers, if QICs can improve the way in which we adjust for differences in risk among hospitals, then they might help groups more accurately determine

which hospitals to model their practices after. This is particularly important for efforts that seek to improve quality in a specific condition, operation, or injury since hospitals that perform well in one domain (e.g., penetrating trauma) may not perform equally well in others (e.g., TBI). Second, if QICs enable clinicians to play a larger role in developing the risk-adjustment models that will be applied to their data, then they may also increase the face validity and credibility of the ultimate results. Multiple studies have shown that the more clinicians believe that the data they receive, the more likely they will be to use these data to improve care. If QICs can increase "buy-in" to the benchmarking process, then they may be critical in helping to break the divide between clinician and non-clinician quality staff that plagued our previous experience with CQI.

Chapter 6 focuses on the related issue of selecting an appropriate intervention prior to launching a large-scale quality effort. Here, I argue that QICs may allow hospitals to quickly determine whether interventions from another setting, such as a national set of evidence-based guidelines, would produce the same benefits in their local environment. Unlike individual hospitals, which are limited by their own sample of patients and their own set of clinical practices, QICs can paint a more accurate and reliable picture of how care is delivered across the entire geographic region. This picture--really the clinical data collected and analyzed by the group--can then be used to determine whether the same patterns of care (and context) that allowed an intervention to be successful elsewhere also exist locally.

Within LACTC, we used this idea of the QIC as a type of multi-institutional PDSA cycle to help us decide whether or not to launch a countywide intervention to increase the use of ICP monitoring in TBI, a practice that is currently recommended on the national level by the Brain Trauma Foundation. Using pooled data from all 14 LA County trauma centers, we tested whether centers that had higher rates of ICP monitoring also had lower risk-adjusted mortality rates--a

quick and easy test of what we might expect from our proposed intervention. Somewhat surprisingly, we found no association between compliance with the guidelines and mortality and subsequently concluded that increasing compliance rates would not be an appropriate clinical target for our group. In this way, the QIC approach not only saved time and resources, but allowed the group, then in its earliest stages, to build confidence and trust by focusing on interventions that were more likely to be a success.

Without well-designed clinical trials comparing hospitals participating in QICs to similar hospitals attempting similar quality improvement programs on their own, we may never be able to determine if QICs are more effective than single-institution CQI. However, by understanding the mechanisms behind QICs and establishing their place in the larger armamentarium of quality improvement tools, my original research aims to shed some light on what benefits QICs offer, when the approach might be useful to hospitals, where it should and should not be applied, and, perhaps most importantly, how it can be modified and improved to best meet the needs of its participants.

Chapter 4 Inside the "Black Box": Developing a Conceptual Model for Quality Improvement Collaboratives

Abstract

Context: Quality improvement collaboratives (QICs) have been promoted as a method of improving hospital quality since the late 1980s. Although there is some evidence to support the use of QICs, many aspects of the approach--particularly how QICs work and why they vary in their effectiveness--remain poorly understood. We sought to develop a conceptual model for QICs in order to understand what benefits they may offer as a quality improvement tool as well as when and where this tool may be best applied.

Methods: We combined a narrative review of the core literature on QICs and a series of 13 semi-structured interviews with individuals who helped to develop, manage, or evaluate QICs. Based on our findings, we constructed a conceptual model to explain how a hospital's decision to participate in a QIC might affect patient-level health outcomes.

Findings: We identified eight unique benefits of the QIC approach: learning by comparison, transfer of (experiential) learning, geometry of testing, outsider perspective, social support and motivation, technical support, accountability, and authority. Success of the QIC as a whole is driven by inter-organizational factors, namely trusted relationships, open sharing of data and experiences, and a shared commitment to learning. However, the success of individual hospitals is dependent on intra-organizational factors, such as organizational readiness for change and the extent to which QIC benefits align with and satisfy hospitals' quality needs.

Conclusions: QICs offer several unique advantages over single-institution quality improvement efforts. However, their success is often limited by poor group dynamics or hospitals' inability to translate the QIC experience into changes to clinical practice. Future QICs may benefit from a more active and individualized approach to collaboration in which hospitals' readiness for change is formally measured and collaborative activities are chosen based on their ability to build trust, promote sharing, and encourage a commitment to learning.

Keywords: quality improvement, collaboratives, implementation science, improvement methodologies, hospital quality

Introduction

Under growing pressure from patients, payers, and government regulators, hospitals throughout the United States are racing to understand and improve the quality of care they deliver. Countless articles over the last two decades have documented substantial gaps between the care our healthcare system could deliver and the care patients actually receive.[15,175,176] Left unchecked, these gaps have resulted in a system that produces suboptimal clinical outcomes[177,178] and, perhaps unsurprisingly, leaves patients only moderately satisfied with their healthcare experiences.[179]

Few publications speak to both the gravity and pervasiveness of our healthcare system's quality problems like the Institute of Medicine's landmark 2001 report *Crossing the Quality Chasm*,[180] which, in conjunction with reports from the National Roundtable on Health Care Quality,[181] the Advisory Commission on Consumer Protection and Quality,[182] and the RAND Corporation,[175] helped to bring the issue of healthcare quality into the national spotlight. Yet, even if these and other high-profile publications have succeeded in generating a national consensus on the need for improve care, no similar agreement has been reached with respect to the optimal strategy for achieving our quality goals.

Since the advent of quality assurance in the 1970s, experts have touted a wide variety of improvement tools with varying degrees of success. Utilization and peer review were quickly met with distain from physicians.[183] Guideline-based interventions--even those using evidence-based guidelines--often suffered from poor compliance, due, in part, to their attempts to change providers' behavior without simultaneous re-shaping the systems in which they work. More ambitious efforts, such as continuous quality improvement/total quality management, focused on the structured evaluation of hospital processes and on workforces that were empowered to

identify and quickly resolve quality problems.[38] However, several evaluations of these and other offshoots of industrial quality management have primarily demonstrated changes to attitudes and perceptions of quality rather than to actual clinical care or patient outcomes.[34,35]

One strategy that has gained popularity in recent years is the quality improvement collaborative (QIC). QICs build on the basic principles of industrial quality management, but shift the focus from a single hospital trying to understand its own care processes to groups of hospitals working together to find common solutions to a single, shared quality problem.[62] Although the specifics of QICs can vary, the original "Breakthrough Series" model developed by the Institute for Healthcare Improvement envisioned teams from each participating hospital--typically consisting of at least one clinician, one quality manager, and one member of senior leadership--meeting with members of a central organizing committee over a 12-month period.[58,60]

Under this model, the entire group meets in-person at the beginning of the collaborative process to review the current evidence on its chosen topic and to compare risk-adjusted clinical outcomes. Many QICs also employ quality improvement coaches who use this first in-person meeting to train team members in aspects of quality measurement that will be useful for establishing baselines and tracking changes over time (e.g., Plan-Do-Study-Act cycles). Following this first in-person meeting (often referred to as a "Learning Session"), the QIC moves into its first "Action Period," where teams return to their hospitals and begin to implement best practices that were identified through group discussion. Most QICs repeat this Learning Session-Action Period process at least three times to allow individuals the opportunity to share their experiences and to allow the group as a whole to build upon individual successes and failures. Proponents of the QIC approach believe that this open sharing of both data and personal

experiences allows hospitals to more accurately and rapidly characterize their own gaps in quality, identify regional best practices, and implement interventions based upon these practices at their own home institution.[11]

Yet, for all of its theoretical benefits, the QIC approach has produced decidedly mixed results. Two systematic reviews and a handful of consensus papers provide isolated examples of QICs significantly improving clinical outcomes,[62,75-80,85] but also highlight several notable failures.[90] Based on these findings, one systematic review concludes that "the evidence [for QICs] is positive but limited"[63(p1498)] while the other states that "the existing research literature provides only limited support for the overall effectiveness of QICs in improving patient outcomes."[64(p388)]

As with other quality improvement interventions,[174] the lack of either conclusive evidence or even a compelling theory to explain this variable success has left both improvement experts and hospital CEOs struggling to decide whether QICs are worth the substantial investment they often require.[31] Part of this difficulty in determining QICs' value as a quality improvement tool is undoubtedly methodological. Despite the existence of dozens of QICs since the late 1980s, there remains no well-defined or widely-held measure for QIC success. And, while early evaluations may have correctly chosen to focus on qualitative experiences and generic summary scores in order to promote QIC development,[97,98,102] the lack of generalizable data on clinical outcomes limits our ability to perform even the most basic comparative effectiveness research. Moreover, since participation in a QIC is typically voluntary, hospitals that place a higher value on quality improvement or have more experience implementing improvement programs may also be more likely to join QICs, thereby biasing traditional statistical comparisons.

However, another, perhaps larger, source of our difficulty in determining the value of QICs rests in how we have been framing the question. Previous research has focused almost exclusively on trying to determine *whether* QICs work, which neglects important differences both among QICs and among hospitals participating in the same collaborative process.[106] Asking the question in this manner is analogous to asking whether chemotherapy works. For certain patients and certain conditions, it may. But before a physician prescribes it, she needs to understand both how the medications work and whether the regimen she has selected is appropriate for her patient's tumor biology. These two areas, in particular--*how* QICs work and *when* they are more effective--are where the current literature on QICs is at its thinnest.

Our study attempts to establish a unified framework for QICs by integrating data from the published literature with years of experience establishing, running, and evaluating QICs. In developing this framework, we sought to answer two basic questions. First, through what mechanisms does participation in a QIC affect patient-level health outcomes? Put another way, what are the marginal benefits of participating in a QIC that cannot be gained from single-institution quality improvement work? And, second, why does the QIC method appear to work in some settings, but not others? More specifically, what is it about the QICs and the hospitals within QICs that do succeed sets them apart from those QICs and those hospitals that fail?

In answering these questions, we hope to address two specific aims: 1) to improve our collective understanding of QICs as quality improvement tools, particularly when and where they are most effective; and 2) to arrive at a series of practical recommendations both for hospitals deciding whether or not to join a QIC and for QICs looking to adapt their efforts to better match their participants' quality needs.

Methods

Given the variety of materials published on QICs and the lack of an established framework for exploring QIC function, we chose to combine two research methods: literature review and qualitative interviews. Our approach was meant to approximate the *meta-narrative review* described by Greenhalgh and colleagues,[184] with one key methodological difference. Instead of relying solely on seminal articles and published models to describe the central "storyline" of QICs, as Greenhalgh and colleagues describe, we also interviewed key individuals who either contributed to or were influenced by this storyline. By combining methods, we hoped to complement the structure and scope of a formal literature review with the richness and depth of qualitative research.[185] Others have also used a similar approach in order to gain a more complete and practical understanding of a given topic.[34,109]

We explicitly chose to explore how participation in a QIC affects patient-level health outcomes rather than hospital-level process changes for two reasons. First, since QICs are designed to influence both the choice of intervention and its implementation, focusing only on how participation affects intermediate process measures (e.g., compliance with a particular evidence-based care bundle) may not fully explain how QICs affect patient health. Second, although certain components of quality interventions, such as their implementation, may be tracked most effectively using process measures,[12] we believe that the common end-goal of all quality improvement activities should be better health outcomes. If QICs improve compliance, but fail to actually improve clinical outcomes, we would still question their effectiveness as a quality improvement tool. Our approach allows us to not only understand how QICs affect hospital processes, but to identify any barriers that may exist in translating these process improvements into better patient health.

We began by generating a list of the seminal articles on QICs, starting with the initial descriptions of the QIC approach[11,60,61] and then using a combination of reference mining (i.e., a review of cited materials) and reverse reference mining (i.e., a review of other articles citing this article) to add to this list. After an initial list was established, we collected the pertinent Medical Subject Headings from each article and used them to search PubMed for any core articles that we may have missed (Appendices 4.1 and 4.2). While we acknowledge that our search procedure has limitations (namely the possibility that certain articles may have been overlooked in our analysis), it did identify both of the most widely cited systematic reviews on the topic.[64,108] Moreover, we expected any gaps in the literature--whether they be due to our search procedure or a lack of published materials--to be filled in by our key informant interviews.

Given our interest in explaining the variability in QIC success, we paid special attention to articles that either identified specific QIC components and processes or evaluated the relationship between these components and processes and QIC success. We defined *QIC components* as structural characteristics of the group (e.g., number of members) or of hospitals within the group (e.g., previous quality improvement experience). *QIC processes* were specific activities or events that took place as a part of the collaborative experience (e.g., in-person learning sessions). As *QIC success* itself was variably defined, we recorded the metric each study chose, but did not exclude studies based on their selection. Two authors then independently extracted a list of QIC components and processes from the articles we had identified and recorded whether or not the association between each component or process and the authors' chosen measure of QIC success was formally tested. If so, they also recorded whether the association was consistently positive (i.e., a higher likelihood of QIC success), consistently negative (i.e., a lower likelihood), consistently zero (i.e., no change in likelihood), or mixed

across the studies. From this initial list, the same two authors grouped items into common themes and then compiled all of pertinent information from our literature review into a table that could be shared with our interviewees (Appendix 4.3).

After completing our literature review, we performed a series of semi-structured interviews with 13 experts in order to supplement our understanding of the QIC approach. Experts were chosen based on their personal experience with either designing, managing, or evaluating QICs (Table 4.1). All individuals were identified either through group discussion or by the suggestion of other experts (i.e., "snowball sampling"). Interviews were loosely based on a series of questions in six subject areas (Table 4.2), but were allowed to take whatever form best suited the knowledge and experience of the particular interviewee. Prior to the interview, interviewees were provided with the initial list of QIC components and processes that we had compiled from the literature and a short description of the Consolidated Framework for Implementation Research (CFIR), which we intended to use as a guiding framework in our model. These materials were used to help frame the discussion on how QICs might affect hospitals and as the basis for our specific questions on QIC components and processes. All interviews occurred between February-March 2016, lasted between 30-90 minutes, and took place either in person or via telephone, depending on the location and preference of the interviewee. When possible, interviews were taped and transcribed. After completing all 13 interviews, we performed a two-stage analysis on the interview transcripts by first dividing quotations into three categories based on our study questions and then grouping quotations within categories by common themes.[186] The Office of Human Research Protection Program at the University of California, Los Angeles reviewed our study and determined that this work did not constitute human subjects research.

Key Findings

QIC Components, Processes, and Group Success

Literature

We identified 14 articles reporting on QIC components or processes.[40,62,64,108,109,117,187-193] A description of study characteristics is provided in Table 4.3. Two of the articles we identified actually updated findings from another article in the group: Dückers et al. 2008 developed an instrument for measuring QIC components that was then tested in Dückers et al. 2009 and Campbell 2009 provided a narrative description of key QIC components that was supported with data in Campbell et al. 2010.

The 12 unique articles used a variety of methods, including surveys (four studies), interviews (two), systematic review (two), case studies (two), and consensus panels (two). Four studies specifically focused on QICs from the Institute for Healthcare Improvement's Breakthrough Series. The remainder either included QICs based on a variety of models (five studies) or focused on one specific QIC program (two studies: one described a series of multi-level QICs in the Netherlands [i.e., each hospital could be in multiple collaborative programs] while the other focused on a regional surgical collaborative in Michigan). The vast majority of studies used subjective measures of QIC success and did not perform statistical testing to evaluate the association between QIC components or processes and QIC success.

Table 4.4 presents the 46 QIC components and processes identified by our literature review, which we grouped into eight categories based on common themes. Overall, there was no clear component, process, or even category of components and processes that was routinely associated with QIC success. Of the 46 components and processes we identified, only six had a

Table 4.1: Names and titles of QIC experts

Name	Title
Paul Batalden, MD	Professor Emeritus, The Dartmouth Institute for Health Policy and Clinical Practice, The Geisel School of Medicine at Dartmouth
Darrell Campbell, Jr, MD	Chief Medical Officer, University of Michigan Health System; Professor of Surgery, University of Michigan School of Medicine
Robert Cherry, MD	Chief Medical and Quality Officer, UCLA Health System
David Flum, MD, MPH	Director, Surgical Outcomes Research Center; Associate Chair for Research and Professor of Surgery, University of Washington School of Medicine; Founder, Surgical Care and Outcomes Assessment Program
Lynn Garofalo-Wright, DPPD, MHA	Managing Director, Healthcare Performance Improvement, Kaiser Permanente – Southern California
Donald Goldmann, MD	Chief Medical and Scientific Officer, Institute for Healthcare Improvement
Oscar Guillamondegui, MD	Vice Chairman for Surgical Quality, Safety, and Professionalism, Vanderbilt University Medical Center; Professor of Surgery, Vanderbilt University School of Medicine; Chair, Tennessee Surgical Quality Collaborative
Julie Johnson, MSPH, PhD	Professor of Surgery, Feinberg School of Medicine (Northwestern University); Associate Director for Evaluation, Illinois Surgical Quality Improvement Collaborative
Kedar Mate, MD	Senior Vice President for Research and Development, Institute for Healthcare Improvement
John Øvretveit, PhD	Director of Research and Professor of Health Innovation Implementation and Evaluation, Karolinska Institutet
Paul Plsek, MS	President, Paul E. Plsek & Associates, Inc.
Janet Rimicci, RN, MSN	Executive Director for Quality and Patient Safety, UCLA Health System
Lucy Savitz, MBA, PhD	Director of Research and Education, Institute for Health Care Delivery Research, Intermountain Healthcare; Research Professor of Clinical Epidemiology, University of Utah School of Medicine

consistently positive association and the majority of these six were only consistent across one or two articles.

Governance. Four articles suggested that QIC governance was important to success, although none specifically tested this hypothesis. In general, QICs that united around a common mission, instituted clear goals, and established rules of behavior for the group were thought to be more likely to be successful.

The issue of external funding (i.e., from a sponsoring organization rather than the hospitals themselves) appeared in several articles; however, there was no clear consensus as to whether this was a positive or a negative influence. Multiple articles from the Michigan Surgical Quality Collaborative claim that their "pay for participation" framework, in which an external sponsor pays every hospital the same stipend for participating rather than paying hospitals based on their relative performance, allows smaller, less financially solvent medical centers to participate and encourages free sharing of data and improvement experiences.[121,169] Others have suggested that forcing hospitals to pay their own way creates "skin in the game" and motivates senior leadership to stay involved with the process. No article actually compared success between QICs with and without external funding.

Topic. Three articles listed the clinical relevance of the topic to the team and to the front-line providers in the organization as key to QIC success. This was generally based on the idea that front-line providers would be more invested in the collaborative process and more willing to implement suggested changes if they felt that the work would impact day-to-day patient care. However, the one article that tested this idea found no significant association with QIC success.[108] Two articles suggested that the complexity of the topic might affect success, with more complex topics or topics requiring more organizational investment being less likely to

Table 4.2: Topic Areas for Semi-Structured Interviews

1. What benefits do QICs offer above and beyond what hospitals can do on their own? How might a hospital decide between participating in a QIC and pursuing its own single-institution quality improvement programs?
2. Which QIC components or processes drive QIC success? Please review and comment on our initial list of QIC components and processes. What is missing or redundant? Which are the most important to success?
3. In thinking about the CFIR framework (intervention, implementation process, individuals, internal setting, external setting), which hospital domains are affected by participation in a QIC? To what extent?
4. Can you identify any particular components or processes that affect particular domains? For example, if the credibility of the experts is important, does that affect the individuals more than the internal setting?
5. Are there hospitals that seem to do particularly well or poorly in a QIC? What about those hospitals makes them good or bad candidates for QICs? What characteristics or processes do hospital need to have or do before joining a QIC to ensure their success?
6. Are there clinical problems that seem to particularly benefit from or not benefit from the QIC process? What about those problems makes them good or bad candidates for QICs?

produce success; [109,194] however, neither article tested this hypothesis. The only concept that did seem to be universally supported by the literature was that successful QICs chose a topic that was of strategic importance to hospitals in the group. Both studies that formally tested this concept found a positive association for hospitals within QICs (i.e., hospitals that reported a higher strategic interest in the topic were more likely to demonstrate larger improvements to quality markers)[187,190], although no study attempted to compare levels of strategic importance among QICs.

Experts. Several articles listed the credibility, helpfulness, and knowledge of experts as important predictors of QIC success. When experts were viewed as more credible by QIC

participants--particularly clinicians--this was thought to increase the likelihood that teams would adopt their suggested changes; however, this idea was not formally tested. QICs in which participants rated experts as helpful were consistently more likely to succeed. QICs in which participants rated experts as knowledge were typically more likely to succeed; however, this trend did not carry across all studies.

Team. A wide range of team attributes and aspects of team function were thought to be predictive of QIC success. Here, we defined *teams* as the individuals from each institution that participated in QIC events in order to separate them from others in the organization. Only one component, team size, was found to have a consistent association with QIC success, although this association was only tested by one article in one systematic review.[108] Interestingly, in that article, team size appeared to have a non-linear association with QIC success, with additional members improving performance up to a point, but then detracting from team function. Stronger leadership, previous experience working together as a team, and maintaining an open and non-punitive climate were also thought to be important; however, there was not a consistent association between any of these factors and QIC success across all studies.

Organization. Characteristics of the hospitals themselves were among the most cited and most tested QIC components. Only two of these characteristics--time and resources--were easily quantifiable; the remainder dealt with the values, opinions, and attitudes of the organization more generally. Organizational culture was listed separately by only one article, however, several of other components are undoubtedly influenced by culture. Implementation climate, which combines the importance senior leadership attaches to quality improvement with the organization's ability to actually carry out change, was listed in two articles, but not consistently associated with QIC success. Interestingly, several articles suggested that an organization's

Table 4.3: Description of studies listing QIC components or processes

Authors, year	Study design/ source of data	Number of QICs	Types of QIC	Country	Measure of success	Associations between QIC component/process and QIC success	
						Statistical testing?	Proportion of significant associations
Plsek 1999[40]	Case studies	Not stated	BTS, VON, others	USA	Expert opinion (not explicitly defined)	No	-
Øvretveit et al. 2002[62]	Consensus meeting of QIC experts	Not stated	Various	North America, Sweden, UK	Expert opinion (not explicitly defined)	No	-
Wilson et al. 2003[109]	Semi-structured interviews with QIC leaders	15	BTS only	Australia, France, the Netherlands, Norway, Sweden, UK, USA	Expert opinion (not explicitly defined)	No	-
Mills and Weeks 2004[187]	Pre-post survey of QIC participants	5	BTS only	USA	≥20% improvement in at least one outcome measure for at least two months	Yes	5/41
Ayers et al. 2005[188]	Semi-structured interviews with QIC participants and leaders	10	Various	Sweden, UK, USA	Expert opinion (based on the utilization of evidence-based best practices, the inclusion of clinical improvement experts, and the demonstration of "improved care and data-driven outcomes")	No	-

Dückers et al. 2008,[189] Dückers et al. 2009[190]	Cross-sectional survey of QIC participants	18	Multi-level, hospital-based	The Netherlands	Self-reported success, project-specific process and outcome metrics	Yes	2/3 for self-reported success, 0/3 for performance indicators
Nembhard 2009[192]	Cross-sectional survey of QIC participants	4	BTS only	USA	Self-reported helpfulness, external rating of improvement	Yes	7/12 for helpfulness, 5/12 for improvement
Campbell 2009,[191] Campbell et al. 2010[117]	Case study	1	Regional	USA	Expert opinion (not explicitly defined)	No	-
Schouten et al. 2010[193]	Cross-sectional survey of QIC participants	2	BTS only	USA	Expert opinion (based on "sufficient expert panel support," "effective multiprofessional teamwork," "appropriate use of the improvement model," and "helpful collaborative processes")	No	-
Hulscher et al. 2013[108]	Systematic review	26	Various	Not stated	Project-specific process or outcome measure (13 studies), self-reported success (10 studies)	Yes	59/200
Nadeem et al. 2013[64]	Systematic review	13	BTS, CCM, VON, others	Not stated	Provider-level outcomes (11 studies), patient-level outcomes (4 studies)	Yes	0/14
Øvretveit 2013[194]	Consensus meeting of QIC experts	Not stated	Various	USA, UK, "Nordic countries"	Expert opinion (not explicitly defined)	No	-

ability to translate work from the QIC to others in the organization was important to actually improving health outcomes; however, there was no consistent association with QIC success.

Characteristics of and relationships among teams. Of the six components related to how teams were constructed or how they related to one another, only one--communication among teams--was consistently associated with QIC success. QICs in which teams had better communication and stronger social networks were both more likely to report that the QIC approach was helpful and more likely to be rated as having made significant improvements by external organizers.[192] Baseline performance was associated with QIC in some studies, but not others. The remaining items were not formally tested, except for voluntary versus mandatory participation, which was not associated with success in two separate studies.[64,108] However, since the vast majority of QICs rely on voluntary participation, there may be other important differences between voluntary and mandatory QICs, such as the reasons for forming, the underlying model for collaboration, or the type of hospitals involved.

Improvement activities. We identified seven improvement activities and seven collaborative activities from the included articles. We defined *improvement activities* as actions that teams performed to either understand or correct quality problems. These ranged from reviewing the literature on a given topic and collecting data to actually developing and carrying out interventions. Only two improvement activities--the change package and literature reviews--had a consistently positive association with QIC success; however, each was only tested in one article. The remaining activities either demonstrated mixed associations or no association with success. Interestingly, participating in novel data collection efforts was listed by four articles, but was not associated with QIC success. Since data and data sharing are defining elements of the collaborative process, we interpret this finding to mean that collecting additional data

specifically for the QIC rather than as a part of clinical care or other quality reporting programs may not be necessary to improve quality. However, further research is needed.

Collaborative activities. We defined *collaborative activities* as actions that either facilitated or relied upon relationships among hospitals and, by definition, could not occur during single-institution quality efforts. This category included both specific activities (e.g., outcome benchmarking, site visits) as well as a variety of methods in which hospitals could communicate with one another (e.g., face-to-face, through reports, electronically). None of the listed activities was consistently associated with QIC success, although several were not formally tested.

A related study by Nembhard (2012), which was not included in our sample because it did not specifically list QIC components or processes, did find a significant association between participation in inter-organization learning activities (e.g., conference calls, collaborative extranet, listserv discussion) and hospital success within QICs.[72] It is unclear, however, whether participating in these types of activities or being the type of hospital that is willing to participate in these types of activities drove this finding.

Interviews

In contrast to the published literature, interviewees tended to focus on the overall collaborative atmosphere rather than specific components and processes. Of the eight categories we identified from our literature review, interviewees spent the most time discussing the *characteristics of and relationships among teams*, especially the way teams interacted with one another and their beliefs about the purpose of the group. Four concepts, in particular, were repeated by multiple interviewees:

Table 4.4: QIC components and processes

Component	Definition	Tested?	Direction of the association	Source
Governance				
Clear and shared mission	Whether the collaborative collectively agrees to a specific set of objectives than can be explained by any member of the team	N		188,191,194
Rules for appropriate data use and to protect confidentiality	Whether the collaborative creates a written or oral agreement regarding how data generated during the process are to be used and not used (e.g., whether or not unblinded reports will be used, whether or not results can be used for marketing purposes)	N		188,191
Funding	Whether the collaborative receives external funds for administration (i.e., from a government or sponsoring organization), fees from the participating organizations, or no additional funds	N		117,188,191
Tangible and achievable goals	Whether the collaborative establishes clinical or administrative goals that can be measured over time and met within the planned timeframe	N		188,194
Topic				
Clinical relevance	Extent to which clinical members of the team and front line providers in the organization identify the topic as an important quality problem	Y	0	108,109,188
Complexity	Whether current evidence exists to direct improvement efforts and the extent to which interventions will require change in multiple organizational microsystems, particular those that are not directly represented by the team	N		109,194
Strategic importance to organizations	Extent to which senior leaders believe improving quality in the selected clinical area is essential to fulfilling the organization's mission	Y	+	109,187,189,190
Experts				

Credibility of experts	Extent to which teams trust the knowledge and experience presented by the experts	N		109,193
Helpfulness of experts	Ability of experts to communicate new information and strategies or to help teams adapt new information or strategies to their own institution	Y	+	108,109,189,190,192,193
Knowledge of experts	Extent to which experts contributed either new clinical information or new ideas and strategies for improvement	Y	+/-	108,189,190,193
Team				
Clear team roles	Extent to which each team members knows what he or she is required to do during the collaborative and the extent to which these roles overlap	N		188-190,193,194
Multidisciplinary membership	Whether teams include members from each essential organizational microsystem, typically a clinician, a senior leader, and a project manager or QI specialist	N		109,187,188
Team climate	Extent to which team members understand each other's strengths and weakness, feel respected, and speak their mind without fear of reprisal or rebuke	Y	+/-	108,109,187,189,190
Team experience	Extent to which team members have worked together previously or worked on other team-based projects	Y	+/-	108,187
Team leadership	Ability of at least one member of the team to take charge, resolve disputes, and hold the team accountable for meeting its goals	Y	+/-	108,187,193
Team size	Number of members on the team	Y	+	108,109,188
Shared goals	Extent to which team members agree as to why they are participating in the collaborative and what they can gain from the experience	N		187,193
QI experience	Extent to which team members had participated in previous QI programs or interventions	Y	+/-	108,109

Organization				
Culture	Set of beliefs, standards, and practices that govern how employees in the organization behave and relate to their work	Y	0	[108]
Dissemination outside collaborative team	Extent to which team members shared the results and knowledge generated by the collaborative with other members of the organization	Y	+/-	[64,108,187,194]
Frontline support outside of collaborative team	Degree to which clinicians (i.e., physicians, nurses, clinical specialists) agree with the work of the collaborative team	Y	+/-	[108,187,192]
Implementation climate	Extent to which and speed with which specifics changes can be implemented within an organization (reflects both the organization's commitment to QI and its readiness to change)	Y	+/-	[108,109]
Leadership support outside of collaborative team	Extent to which the team and other members of the organization feel senior leadership provides them with the necessary support to achieve their intended goals	Y	+/-	[64,108,189,190]
Resources	Whether sufficient financial and organizational resources were provided to allow team members to participate in the collaborative and to implement the results of the collaborative	Y	+/-	[108,109,187,193]
Time	Whether sufficient time was made available to allow team members to participate in the collaborative and to implement the results of the collaborative	Y	0	[108,187,193]
Characteristics of and relationships among teams				
Baseline performance	Degree to which organizations' initial clinical outcomes differ from the targets and goals of the collaborative	Y	+/-	[108]
Collaborative size	Number of teams in the collaborative	N		[188]
Communication among teams	Extent to which collaborative develops or matures social networks among teams that allow for formal and information sharing of information or practices	Y	+	[109,192,193]

Team identity	Extent to which teams believe the collaborative will address aspects of the topic or implementation than may be unique to their organization	N		[188]
Trust among teams	Willingness of teams to freely share their own data and experiences with other teams in the collaborative	N		[188,191,193]
Voluntary vs. mandatory participation	Whether teams or organizations chose to be a part of the collaborative vs. being placed into the collaborative by an external body (e.g., government, hospital system)	Y	0	[64,108]
Improvement activities				
Change package	Set of interventions developed during the collaborative that reflects its collective understanding of the topic and can be used as a template for making local changes at each participant hospital	Y	+	[192]
Data collection	Extent to which teams are required to collect new data on their processes or outcomes as a part of the collaborative process	Y	0	[64,108,187,191]
Focus on measurement	Extent to which collaborative activities are focused around establishing metrics and using those metrics to drive and measure changes in care	Y	+/-	[108,193,194]
Goal setting	Extent to which teams are required to develop process or outcome measures that can be used to track performance throughout the collaborative process	Y	+/-	[108]
Literature reviews	Systematic collection of published literature on the clinical topic and dissemination of that information to collaborative teams	Y	+	[192]
Pre-work	Whether organizations complete of a specific set of activities prior to the first collaborative meeting, typically an internal review of institutional data and practices or readings on evidence-based best practices related to the clinical topic	N		[194]

Training in QI methodology	Extent to which teams are taught the principles of modern QI theory or trained to employ specific QI techniques, such as PDSA cycles, root-cause analyses, or fishbone diagrams	Y	+/-	64,109,188,192
Use of technology to manage and present data	Extent to which either the collaborative or teams within the collaborative use technology to collect, store, and compare data	Y	0	108,187,188
Collaborative activities				
Communication outside of meetings	Electronic (email, listserv, blog, website) or telephonic communication among teams to discuss collaborative activities or progress	Y	+/-	64,108,188,192,193
Face-to-face meetings	In-person learning sessions, typically held 2-3 times during the collaborative	N		109,188,193
Outcome benchmarking	Whether teams explicitly compare their clinical outcomes for a given condition to those of other teams in the collaborative	Y	0	64,188
Process benchmarking	Whether teams explicitly compare their clinical processes for a given condition to those of other teams in the collaborative	N		188
Sharing information and progress through reports	Written or electronic documents reviewing the current status and progress of all collaborative teams based on a series of agreed upon metrics	Y	0	64,109,187,192-194
Site visits	In-person meetings held on location at one or more member sites with the goal of witnessing and understanding differences in care process	Y	0	188,192
Social events	In-person meetings held outside the collaborative with the goal of developing and maturing social networks	N		188

Trusted relationships. Overall, the building of trusted relationships among participants was seen as the most important determinant of QIC success, primarily because this facilitated the open sharing of data and improvement experiences. Building trust was felt to be particularly important at the beginning of the collaborative when most teams entered as "strangers." Although teams themselves played a role in building trust (e.g., one QIC actually had a buddy system for new members to the group), most interviewees felt that trust was primarily the responsibility of QIC organizers. One interviewee described her role as an organizer by saying:

> In every collaborative I've ever led, I think the secret sauce is having trusted relationships…you[, as the organizer, have to] demonstrate value [so that] people learn, "this is how it works. I can trust you with things."

Several factors were seen as important to building and maintain trust. Interviewees felt that organizers had to create a "safe space" where teams could share their thoughts and experiences without judgment from others in the group or from others in their organization. One interview commented that successful QICs created an "open dialogue in a safe environment." Another said:

> When there is a real culture of celebration and recognition in collaboratives, those tend to be more successful. I would say, above all else, if a collaborative is tainted by the spirit of reprisal or fear or judgment or punitive action or measurement for the sake of any of those things, the effectiveness of the collaborative plummets.

Creating this safe space often required organizers to lead by example. Some chose to share experiences with quality programs where they themselves had failed while others firmly established rules about proper sharing and cut off statements that they felt were judgmental instead of constructive.

Several interviewees cited legal partnerships and data use agreements as ways to build trust. One said:

> We've established that trusted relationship through our master collaborative agreement so there are no intellectual property issues. There are no issues around commercialization within our data trust. We're sharing incredible data that allows us to take a window and look into what we're doing.

Another focused on the combination of social capital and legal agreements as tools for developing trusting relationships within the group:

> Social events, face-to-face meetings, and agreements on data use and confidentiality are all ways to build trust among the teams so that they can openly and more effectively and efficiently share the details about their practices. What's important is that they share the details of their practices. And are things that people have tried to get the level of trust high enough that you can share openly.

Open sharing of data and experiences. On an operational level, interviewees felt that QICs were successful when the sharing of data and personal experiences became the main focus of group meetings. Not only did the information itself help to guide what the QIC worked on, but the act of sharing data and opening oneself up to scrutiny was seen as a sign of commitment to the group. In that way, a kind of positive feedback loop was formed: building trust led to open sharing of data and open sharing of data led to higher levels of trust and more energetic collaboration. One interviewee said:

> Collaboratives tend to do better when data is really shared…data drives the sharing process and makes it a lot more tangible. And as a corollary to that, transparency tends to create better opportunities for learning and collaboration. In certain situations, people have put up various barriers to transparency. Some of those are logical and some of those aren't. But the more transparent the better, I would say. We tend to be more successful when the data are more transparent in collaboratives.

Another framed the importance of transparency even more simply: "You can't have any secrets if you're going to learn."

While the open sharing of both clinical data and experiential knowledge were felt to be important, each was thought to play a different role in QIC success. Sharing data primarily allows teams to understand their performance relative to others and helps to identify or validate

particular quality problems. Sharing experience, on the other hand, helps teams identify the root causes of quality problems, allows teams to be more efficient in their improvement work, and familiarizes teams with how to overcome common barriers to improvement. For example, sharing data might help a team learn that they have a higher than average surgical site infection rate, but sharing experience teaches team members which interventions have actually reduced infection rates and how to implement these interventions in a hospital that is similar to their own.

Several interviewees mentioned competition among hospitals as a potential detriment to transparency and open sharing. One felt that competition affects QICs prior to formation by limiting the number of organizations that choose to participate (i.e., some hospitals would not join if one of their rivals was already in the group). Others thought that competition could actually be beneficial if it drove organizations to work harder at improving care, but only if transparency and data sharing were maintained. When asked to explain why competition might influence QIC effectiveness, one interviewee responded:

> It's a problem with transparency. That's why the pre-condition of transparency is so critical. We found the people who are in competition often do, in fact, enjoy the collaborative because there is a little edge to it. On the other hand, it does tend to make people a little more secretive, which can be a problem.

Commitment to learning. Interviewees felt that QICs in which hospitals demonstrated a clear commitment to learning were more likely to be a success. This was thought to be the case for at least two reasons. First, hospitals that are committed to learning are more likely to participate in collaborative activities like data sharing. Second, if hospitals believe they can learn from the QIC process, then they will also be more likely to adopt and implement changes from other hospitals. One interviewee commented that:

> Collaboratives tend to work best when people truly embrace the philosophy of everyone having something to teach and everyone having something to learn. When learning is a

> real priority. It's not about discipline and punishment. It's not about getting paid. Learning is the objective.

Another felt that having organizations that were "willing to suspend previously held positions" was essential for QIC success because this curiosity and open-mindedness promoted a culture of learning within the group.

Several interviewees felt that more learning took place when participants were different enough from each other to bring new perspectives and practices to the group. However, this diversity was balanced by the need for enough similarity among hospitals for teams to see the relevance of these new perspectives and for team members to form relationships with individuals holding similar roles in other organizations. One interviewee routinely took time during meetings to break participants up by job description so that clinicians could meet and learn from other clinicians, senior leaders could meet and learn from other senior leaders, and so on. Another said:

> There are virtues to being similar, but there is probably less learning from the heterogeneity. I don't think it works very well if you have a critical access hospital and a Mass General because they are just so different. On the other hand, you don't want five Mass Generals because then your ability to extrapolate the learn is just really really limited. There is a happy medium.

Beyond having some heterogeneity in hospital type, there was no consensus on which components or processes were best at promoting learning. Two interviewees stressed the need for valid data, not just to be sure that teams were designing interventions based on accurate information, but also to lessen skepticism and promote cohesion within the group. Another felt that having efficient, ordered meetings was important to learning because it kept teams engaged and listening when they were not presenting their own results. Still another interviewee thought that maintaining some pressure on hospitals throughout the QIC process was necessary in order to encourage continuous learning:

> We're learning that pace is really important. You have to have pace in the work and kind of an aggressive approach to making change. The critical periods are actually the Action Periods between either the live or the virtual meetings. Most people kind of think that that's a rest period of some type, but it's actually the most important period. So real active coaching and holding people accountable on interim calls and having people share success stories and, hopefully, if they have good culture, failure stories is important.

Topic. Almost every interviewee felt that certain topics were more suited to the QIC approach than others. Better topics tended to meet two criteria: 1) a high degree of interest and relevance to the organization or its representatives; and 2) there exists sufficient content and implementation knowledge for that hospitals to focus on dissemination rather than innovation. Interviewees felt that hospitals were typically interested in "big ticket" items: quality problems that either occur frequently (e.g., nosocomial infections) or impose high costs on the hospital when they do occur (e.g., wrong site surgery). In general, QICs working on clinical topics were thought to be more successful than those working on administrative ones, possibly because clinical topics generate more interest among physicians who have traditionally been reluctant to participate in quality efforts.[35] One interviewee felt that QICs worked well for topics that hospitals "had to work on anyway" due to either internal or external pressure. For example, reducing 30-day hospital readmissions, which are now subject to penalties under Medicare, was suggested as a possible topic for QICs since most hospitals are already working on policies and interventions on their own.

Multiple interviewees felt that QIC only succeed when both the "what" and the "how" of any targeted improvements had already been established. In fact, three interviewees referred, either directly or indirectly, to the original white paper on QICs, which laid out three requirements for potential targets: 1) current practice deviates from best practice, 2) evidence is available to support particular improvements, and 3) at least one hospital has implemented changes that improved care.[58(p3)] QICs that stray from these conditions--either by focusing on

generating knowledge or by choosing a problem that hospitals cannot reliably solve--were seen as less likely to be successful. One interview commented:

> We've gotten away from the original intent of the collaborative and we've used it for a lot of different purposes. Currently, we have a lot of stuff happening in our systems that may or may not match [the original] definition. We might be using the collaborative, for example, to discover new evidence or we might be running a collaborative on a very complicated change that we know works very heterogeneously in different situations for which we have no experience or understanding.

QICs' Influence on Participant Hospitals

Literature

We identified a core group of five to seven studies that is cited by almost every subsequent article on QICs. [11,31,58,60-62,109] Although these studies establish the primary storyline on QIC development, few directly address how QICs influence participant hospitals. Plsek (1997) lists eight benefits of collaboration, but does not specifically state how these benefits affect hospital-level quality improvement processes. Moreover, several of the benefits Plsek lists are not necessarily specific to QICs. For example, "quantified variability in process or outcome," "internal process characterization," "the identification of 'potentially better practices'[11(p90)]," "purposeful replication," and "measured improvement" could all occur as a part of single-institution quality improvement programs.

Kilo (1998) writes that QICs "codif[y] existing knowledge from disparate sources…[build] the internal capacity of participation organizations…monitor appropriate performance measures…[and enable] organizations to use the scientific method to test changes."[60(p12)] Again, however, none of these items appear to be particular to multi-institutional groups. Kilo (1999) does add to this list by suggesting that QICs "[leverage] increased

application knowledge that results when multiple organization or site work together to understand and compare their systems of care."[61(p391)]

Øvretveit and colleagues (2002) provide perhaps the best examples of how QICs affect participant hospitals. They write that QICs "[cut] out much of the investigation work of a traditional quality project…[get] expert support and peer stimulus which might not otherwise be available…[provide] for both professional and organisational development…[and] build interprofessional cooperation."[62(p346)] Mittman (2004) also suggests that QICs "facilitate accurate recognition and diagnosis of quality problems…generate energy and commitment among team members…and provide the team with the knowledge and skills to implement solutions."[31(p899)] Both Institute for Healthcare Improvement (2003) and Wilson and colleagues (2003) discuss features of QICs, but do not explicitly lay out a theory for change.

Interviews

Our interviewees identified eight unique benefits of QICs that they either felt could not be gained from single-institution quality improvement work or were less likely to be realized through such an approach. While some benefits were more commonly cited than others, no single benefit was mentioned by all respondents nor was any benefit consistently reported as the primarily reason hospitals might benefit more from the QIC approach than from single-institution quality improvement efforts.

Learning by comparison. Several interviewees felt that QICs' main benefit was allowing hospitals to compare themselves to others and the knowledge that was generated through this comparison. Both *process benchmarking*, the comparison of hospitals based on how they deliver care, and *outcome benchmarking*, the comparison of hospitals based on what

happens to their patients, were cited as important sources of knowledge. However, each was thought to provide a different benefit to hospitals.

Interviewees thought that outcome benchmarking, which typically occurs first in the QIC process, allowed hospitals to develop a "true understanding of their relative performance" and to "learn where they sit relative to others," which could, in turn, help hospitals identify previously unknown quality problems, validate current quality efforts, and motivate workers to improve care. For example, a hospital might learn that its surgical site infection rate is higher than average and begin working on programs to reduce it. Similarly, a clinician trying to get senior leadership interested in nosocomial pneumonia might gain more support after leaders learn that their hospital has the highest rate in the region. Although outcome benchmarking can occur outside of QICs (e.g., in clinical registries or as a part of governmental quality reporting efforts), interviewees felt that benchmarking within QICs was more directly linked to improvement efforts.

Process benchmarking, on the other hand, is more specific to QICs and can take place either through group discussion or structured site visits. Unlike outcome benchmarking, which can only tell hospitals *that* they are different, process benchmarking helps hospital learn *how* they are different. It also forces hospitals to think about the actions that are responsible for a given outcome and exposes them to potential alternatives to what they are currently doing.[195] One interviewee described the importance of site visits for helping hospitals understand their own internal processes:

> It has been tremendous for our research team to go into the hospitals and see what it is that they are actually doing. Sometimes you see different things from what they say versus what you actually see when you're there.

She went on to say:

> Understanding your processes is vital to being able to drive changes. We know that it doesn't work to just try to pick up something and plop it down in place somewhere else. The only way that people can actually implement changes is to understand their processes and their own work flows and how these changes are going to impact those.

Transfer of (experiential) learning. Two interviewees felt that the main benefit of the QIC approach was its ability to facilitate the transfer of experiential knowledge. Multiple mechanisms already exist for the transfer of technical knowledge in quality improvement. Content knowledge (e.g., which antibiotics are most effective against methicillin-resistant *Staphylococcus aureus* infection) and even improvement knowledge (e.g., how a driver diagram can identify targets for infection control) can be transmitted through research articles, reviews, or didactic sessions. However, the experiential knowledge that individuals and organizations build by participating in quality efforts (e.g., common implementation barriers and how to overcome them) is less easily shared. One interviewee said:

> We don't have a language for discussing the knowledge of how. We've developed methods for standardizing and generalizing the knowledge of what, or the evidentiary base. But the knowledge of how is subject to all kinds of nuances that have to do with the context within which the change is envisioned. Often times, the particular hospital in question doesn't have much insight into what it takes to successfully make change happen in their setting.

Another stated:

> Published articles are written usually at a system level that doesn't get into the granular details that are so important in quality improvement work. So if you bring people together for one of our collaborative meetings and they are talking about the nuts and bolts of how you get people engaged or things they tried that didn't work, those are things you won't necessarily find in the published literature.

This idea of developing a platform for sharing experiential knowledge builds on Perkins and Salomon's work by suggesting that learning--the ability to use experience from one context to influence performance in another[196]--can actually be transferred between people. Perkins and Salomon describe two methods for learning transfer,[197] both of which take place during the QIC

process. The first, *low road transfer*, occurs when hospitals copy practices from another institution and implement them directly. For example, a hospital might take an intervention or checklist that was developed by one hospital in the group and implement it directly at their own institution. The second mode of transfer, *high road transfer*, occurs when hospitals translate another hospital's experiences and use them to inform their own practices. For example, hearing how one hospital struggled to institute a hand-washing protocol might allow another team contemplating a similar project to avoid some of same pitfalls. One interview described this process as "taking away the pieces of the story that apply to you":

> You're listening to them describe the way they did it in their setting. They got everyone with square heads to stand in a line facing East. And you're thinking to yourself, well, we don't have square heads around here. But I could, in fact, get people to face in one direction. So what you've done is, you're heard the square-heads-facing-East stuff and then you've isolated what's important about the story to you in your setting.

Geometry of testing. Most interviewees believed that QICs were an effective method for increasing the efficiency of testing in quality improvement. Under traditional PDSA cycles, an organization implements a single change, studies its effects, and then decides to either continue the change or try a different one based on the results.[17] Because each hospital can only focus on so many tests of change at a time, it can often take several cycles over a considerable period to time to arrive at the optimal solution. QICs, on the other hand, can perform more tests in a given time period by leveraging all of the hospitals in the group. One interviewee said:

> When you think about all the areas where we need to improve quality and safety, your bandwidth is limited as a single entity. In a collaborative, all of us is better than one of us.

Another commented:

> If you're running your own system, you can only run one, two, three--if you're a big system, maybe half a dozen--tests at any given point of in time. In a collaborative learning system, you can learn from ten or twenty or fifty--however many as there are participants in the collaborative. So you have the additional benefit of many, many, many more testers.

Since QICs typically require hospitals to report back on their tests of change, participants are also able to spread the burden of testing across sites without fearing that they will miss out on important findings.

Site-specific testing in QICs can either be deliberate or a result of pre-existing variation in care. For example, a QIC might ask each hospital to choose a different skin prep and use it for all of its colon operations until the next meeting. Or it might compare surgical site infection rates among hospitals in the group (or even among surgeons within the same hospital) based on their choice of skin prep. Performing collective tests across the entire group of hospitals also increases the sample size for statistical comparisons and can impact the external validity of findings if different hospitals that employ the same processes have similar clinical outcomes.

Outsider perspective. Almost every interviewee mentioned that QICs provided hospitals with an external perspective on their own processes and allowed them think outside of their current patterns of care delivery. Futurist Joel Barker coined the term *paradigm paralysis* in reference to situations where innovation is restrained simply because it differs from what people are used to seeing and doing.[198] QICs help to combat paradigm paralysis by exposing participants to different ways of delivering care and then demonstrating through data that many of these alternative strategies result in the same or better clinical outcomes. Interviewees thought that seeing others succeed in different ways could "decrease internal bias" against a particular change as well as help to "overcome the inertia of the status quo and try something new." One interviewee, in particular, felt that QICs could disrupt standard practices not just at the institutional level, but at the individual level as well:

> You might have a clinician on the team that is just a little bit skeptical, that doesn't quite see why we should bother to go about changing. That person may be influenced by someone else in the collaborative. It could be a faculty member or it could be a clinician

> in another team. They meet and they find that they went to the same medical school or they studied under the same folks and they bond. And they go and they talk a little bit. And suddenly that person gets enough energy to make a change because of that conversation that they're having with someone else. Now they get it. Now they want to make it happen.

While this outsider perspective could also be gained through hiring a consultant, interviewees felt that the QIC approach offered a "plurality of experiential voices" and a "much greater experience base on which to operate."

Social support and motivation. Interviewees felt that QICs provided team members with a strong sense of purpose in their work that was useful during the often difficult implementation period. One interviewee said:

> Change of any kind is not an easy process, especially in healthcare where there is a lot of inertia. Feeling supported by a community. Getting reinforcement. Knowing that there are 30 or 40 or 50 other organizations that are trying to do a similar thing builds a sort of support system that allows organizations to make change.

Interviewees believe that the existence of a peer group gives teams a resilience to setbacks and a greater confidence in their ability to make improvements. Several interviewees provided examples of groups taking time to celebrate the successes of individual hospitals or to provide structured counseling to hospitals that were not meeting their objectives. This counseling often went beyond a prescriptive list of suggestions and took on an emotional or motivational tone. Interviewees also suggested that having a peer group helped to maintain a tension for change in organizations that either could not or would not do so on their own.

Technical support. Multiple interviewees commented on QICs' ability to concentrate external resources, energy, and improvement knowledge that might not otherwise be available to certain hospitals. One interviewee, in particular, saw QICs primarily as sources of "temporary change capacity," similar to hiring a consultant, but with more substantial and more varied experience. Another commented that, since "not all delivery systems have the same resources

and capacity," QICs can act as "external quality officers or performance improvement departments" for smaller hospitals. Although this benefit is not specific to QICs, interviewees felt that the structure of QICs allowed them to provide different levels of support since participants could learn from both content experts and peers.

Accountability. Several interviewees suggested that QICs help to hold hospitals accountable for their own performance. Based on the way it was described by interviewees, we divided this concept into two: accountability to other teams in the group and accountability within an organization. *Peer-based accountability* is similar to the idea of peer support in that teams feel like they are letting others in the group down if they do not work as hard as they can on group projects or if they have not improved their own performance as much as others in the group. Although this concept was described by multiple interviewees, a few felt that it did not necessarily pertain to every QIC. One interview commented:

> Learning systems build their own normative description of how accountable they want to be to each other. In some learning systems, the teams hold each other to account with numbers and data and looking carefully at one another's work with a great deal of scrutiny. And in other learning systems it's much more varied and less structured and strict.

Self-accountability, on the other hand, was described as a combination of sunk cost motivation and changes to organizational structure that occur as a result of QIC participation. QICs often entail sizeable commitments in both time and financial resources (e.g., entry fees, travel).[82] While these costs are sunk in the traditional economic sense, interviewees suggested that they may still motivate senior leaders to become more involved in their organization's quality improvement work in order to demonstrate a return on their investment. In addition, the QIC approach requires leaders from multiple organizational microsystems (i.e., senior leadership, clinical, administrative) to work together as a team. Since many quality improvement

projects fail due to a lack of sponsorship, if QICs enable organizations to more effectively communicate across silos, then they may also augment internal support for quality projects and increase the likelihood that improvements are made.

Authority. Two interviewees suggested that QICs may influence hospitals and physicians outside of the group, but in different ways. One felt that QICs provided hospitals with more authority to influence regional and national policy. In particular, this interviewee felt that QICs represented a source of knowledge that could be called upon by state and federal agencies:

> By virtue of the things that we are learning, we can inform policy at a national level with respect to quality improvement. So when I speak to [government organizations], I'm speaking with a little more authority than somebody who comes from a single hospital.

The other thought that QICs' main benefit was "setting professional norms" for hospitals and physicians across the region. He cited changes in attitudes toward public reporting and fewer complex operations being performed at low-volume centers--both targets of their regional QIC--as examples of how QICs can affect hospitals by changing culture rather than through specific interventions.

Sources of Hospital-Level Variability in Success

Literature

No study in the core group of literature on QICs specifically addresses why certain hospitals appear to benefit from QICs while others do not. Two commentaries on the QIC approach, Mittman (2004) and Solberg (2005), come the closest to addressing this point, but only by suggesting that our current methods of evaluating QICs are off base. Mittman (2004) suggests that "determin[ing] whether the [QIC] method is universally 'effective' or 'ineffective' across diverse setting and quality problems" is less important than "develop[ing] insights into the situational factors that facilitate or impede its acceptance, implementation, and effects."[31(p899)]

Solberg (2005) similarly challenges the idea of collectively evaluating the QIC approach, but focuses more on the variety of interventions that have been labeled QICs than or the contextual factors that may drive QIC success. He concludes his commentary by stating:

> We do need more evidence and more creative studies, but those goals will not be facilitated by thinking of QICs as homogeneous. At this stage in their development, it may be more useful to realize that when you've seen one QIC, you've seen one QIC.[106(p199)]

Interviews

Interviewees were able to identify two major reasons why some hospitals improved clinical outcomes in QICs while others did not. As with the benefits of QICs, no single reason for hospital-level variation in success was mentioned by all interviewees nor was any reason clearly dominant.

(Organizational) readiness for change. Nearly every interviewee made reference to the idea that some hospitals entered the QIC process more ready to implement changes than others. Three interviewees specifically referred to the literature on readiness for change as a method for understanding heterogeneity in success, with two going so far as to suggest that future QICs attempt to measure hospitals' readiness for change prior to accepting them into the process.

Although there remains some disagreement within the organizational literature,[199] readiness for change is typically divided into three components: change commitment, change efficacy, and innovation- or domain-specific capacity.[200,201] Our interviewees addressed aspects of each of these components. *Change commitment* refers to the motivation with which an organization pursues change and how much it values improvement. Interviewees typically reference change commitment by discussing whether or not quality improvement was a part of the "identity" of the institution. One interviewee felt that successful organizations would "think

less of themselves if they were not doing quality improvement." Another stated that successful hospitals "believe in improvement, can articulate a clear vision, can articulate clear priories, and can hold people accountable."

Change efficacy, on the other hand, refers to how capable an organization is at actually making changes once it has decided to do so (i.e., does it have a skill necessary to overcome the inertia of the status quo). Here, interviewees referred to hospitals' "capacity to support quality improvement projects" or their "ability to translate findings into action." One interview said:

> Organizations that have a history of being competent and capable at improvement, having the capacity to actually make change, do a lot better than those that just throw a team into the collaborative without any previous experience or infrastructure that can support it.

Interviewees also focused on both whether a hospital had the skills and resources necessary to make changes (e.g., a separate quality improvement team or finances to support physicians who wanted to take time away from clinical care to work on quality projects) and whether a hospital understood its own processes well enough to know where changes need to be made and to recognize when specific changes result in an improvement.[17] One interviewee commented:

> If an institution goes into some kind of a networked activity to try to learn something, if they don't know anything about their own setting or about what works in their setting or why this might be important in their setting, if they don't know any of that stuff then they can't make any meaningful changes.

The final component, *domain-specific capacity*, applies both change commitment and change efficacy to the specific change being made. For example, a hospital might be more interested and more adept at making changes in its transplant surgery program, which it sees as essential to its identity as a hospital, than it is in its pediatrics department. This idea of matching capacity to need and the potential for there to be differences in success based on the topic chosen are explored further in the following section.

Matching supply and demand. Among the most frequent comments made by interviewees was that hospitals succeed in QICs when their activities "aligned with the strategic priorities of the organization." Presumably, this alignment could be achieved in at least two ways: 1) an organization itself could adapt and make collaborative work a core part of its business plan; or 2) a QIC could adapt to meet the interests and needs of its participating organizations. Interviewees were mixed in their beliefs about whether participating in a QIC affected organizations' interest in quality improvement work or their change commitment. Some felt that QICs could affect organizational commitment to quality improvement over the long term *if* they substantially influenced the individuals participating in QIC projects *and* those individuals had enough standing within their organizations to influence others. Most, however, felt that hospitals entered QICs with a fixed level of commitment and that this commitment could only be increased through internal programs or changes in organizational leadership. In fact, many felt that some demonstration of commitment should be a pre-requisite for QIC participation and that hospitals without sufficient commitment would benefit more from other types of quality improvement work (e.g., leadership training or improvement coaches) than from participating in a QIC.

The second pathway--adapting QICs to meet the interests and needs of participant hospitals--was seen as a more viable option. Still, interviewees generally felt that this process does not typically take place in the active sense (i.e., QICs do not actively assess their hospitals' quality needs and then adapt the collaborative process to meet those needs). Instead, interviewees felt that hospitals succeed when their own interests happen to be reflected in what the QIC chooses to work on or when their own needs as an organization happen to be met by the QIC process. For example, a QIC that focused primarily on collecting and tabulating evidence-based

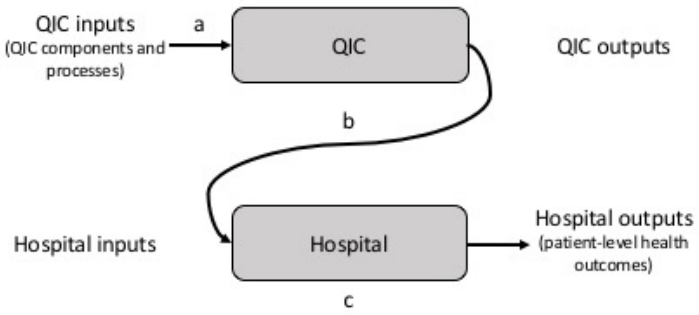
practices might be more beneficial to smaller, regional hospitals, which tend to be less connected to new scientific developments, than to large, academic medical centers. While interviewees generally described the choice of topic as something affecting overall QIC success, several felt that it could play a role in intra-QIC variation as well since some hospitals would be inherently more interested in the topic and might work harder or be more motivated to changes that are needed to improve care.

Components of a Unified Model

Based on the data we collected from our interviews and the published literature, we designed a model for understanding how QICs work and why they may or may not be successful in every setting. Our model has three basic components:

1. **QICs affect patient outcomes by influencing quality improvement efforts at participant hospitals.** Put another way, the impact of QICs on patient health is fully mediated by local changes in either hospital structure or care processes.[12] Although this assumption appears to be obvious, making it explicit allowed us to separate our general model into two constituent pieces: the QIC itself and the individual hospitals participating in the QIC. We then envisioned each piece or level to the model as having its own set of inputs and outputs, as is depicted in Figure 4.1. At the QIC level, shared components and processes act as inputs and influence the successfulness of the collaborative process ("a" in the diagram). This idea aligns with our first category of findings, namely that some QICs are more or less successful that others due, in part, to characteristics of the group or the type of activities it performs.

Figure 4.1: Two-level model for understanding QIC function



Based on these shared inputs, each QIC produces a series of outputs, which serve, in turn, as inputs to hospital-level quality improvement efforts ("b" in the diagram). Again, this design aligns with our second category of findings in that QICs offer at least eight unique benefits that do not occur during single-institution quality improvement work. Although these benefits range from more proximal (e.g., improving organizational accountability) to more distal along the pathway between QIC participation and patient health (e.g., technical support implementing an intervention), all are potentially useful to hospitals working on improving their care.

Finally, hospitals themselves must translate QIC outputs into changes to their structure or processes for QICs to actually affect patient-level health outcomes ("c" in the diagram). This concept reflects interviewees' beliefs that organizational readiness for change and the matching of QIC outputs (supply) to hospital needs (demand) drives hospital-level QIC success.

By envisioning QICs in this way, our model suggests two separate sources of failure. First, a QIC might not produce any benefits for its member hospitals or at least not the

type of benefits that its member hospitals need. Since QIC outputs are driven by QIC inputs, which vary at the QIC-level, this type of failure would be expected to produce *inter-QIC variation* (i.e., one QIC as a whole improving care more than another similar QIC). Although it is purely speculative, this might explain why one QIC targeting childhood asthma resulted in better compliance with evidence-based process measures and asthma-related quality of life[80] while another similar QIC found no differences in either compliance or in patient health.[73]

Since all hospitals participating in the same QIC are exposed to the same collaborative process, we would expect each to achieve the same level of success if QIC outputs were the only factors influencing patient-level health outcomes (i.e., there would be no intra-QIC variation in success). As empirical evidence suggests that this is not the case, our model suggests that hospital-level factors represent a second source of failure, which appears as *intra-QIC variation* (i.e., one hospital in a QIC improving care more than another similar hospital in the group). For example, even within the QIC that did demonstrate improvements in asthma care, not all hospitals achieved the same level of improvement and some did not improve at all.[202]

2. **At their core, QICs are a type of quality improvement intervention.** As such, we believe that they can be best understood within an established framework from implementation science. Several validated models exist.[203-205] We chose to integrate our model into the CFIR framework, both for its general acceptance in the field and its specific focus on explaining differences in success across settings. To the best of our

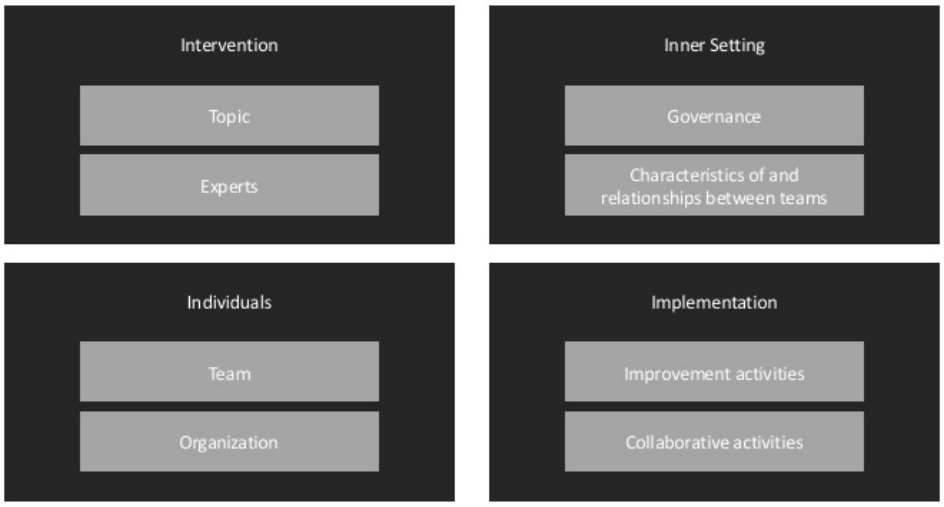
knowledge, CFIR has never been applied specifically to QICs nor has it been validated for understanding differential success and failure among multi-institutional quality improvement programs. Despite this potential limitation, we believe that the five domains identified by CFIR--the *intervention*, the *inner setting*, the *outer setting*, the *individuals*, and the *implementation process*--are sufficiently expansive to be universal for the purposes of understanding quality improvement interventions, regardless of the number of sites involved.

For our purposes, we defined *inner setting* as the characteristics of each member hospital that were not specifically included in either the intervention or the implementation process (e.g., leadership support, organizational culture). *External setting* included the general policies and practices of the U.S. healthcare system that were not specific to any given hospital (e.g., fee-for-service versus capitated reimbursement). We chose not to include external setting as a separate component because we believe its influence is mediated through each of the remaining domains. Individuals, for example, are motivated by financial and non-financial incentives that exist within the current healthcare system. Even the intervention itself may be limited or altered by certain regulations, reporting standards, or insurance contracts.

On a practical level, separating our model into the QIC and the hospitals participating in the QIC allowed us to apply the CFIR framework separately to each level of the model. At the QIC level, we treated the collaborative process as an intervention in its own right and tried to determine which QIC-level domains contribute most to QIC success (e.g.,

how much do the individuals participating in collaborative meetings affect its ability to deliver meaningful QIC outputs?). On this point, neither our literature review nor our interviewees provided us with consistent results. Only a few components and processes were consistently associated with QIC success across studies and interviewees tended to focus on general aspects of collaboration (relationships, sharing, learning) instead of specific characteristics or activities. We believe that these findings--really these lack of findings--have two implications for understanding QIC design. First, rather than focusing on causal connections between specific components and processes and QIC success, future research may benefit from taking a bigger-picture, more domain-based approach. In fact, our eight categories appear to align well with domains in the CFIR framework (Figure 4.2), either of which might be used as a starting point in future work.

Figure 4.2: Crosswalk between categories of QIC components and processes and CFIR domains



Black boxes represent CFIR domains. Grey boxes represent our eight categories of QIC components and processes.

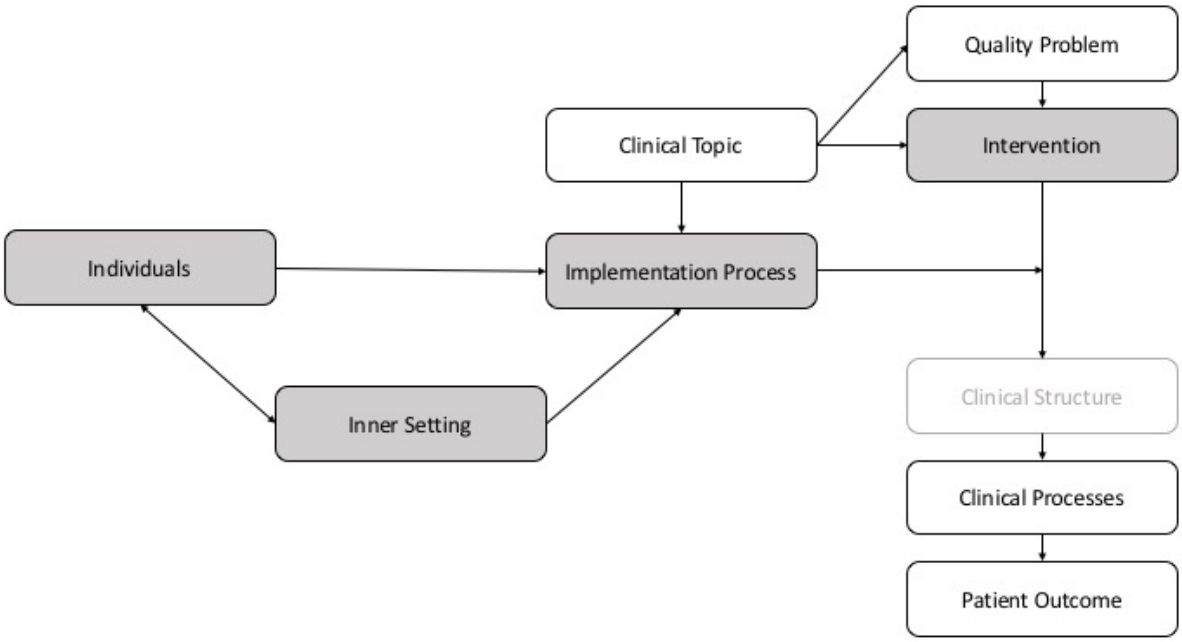
Second, different components or processes may be more or less useful to different QICs. This idea stems from attempting to integrate the literature's granular view of what

influences QIC success and our interviewees' more holistic perspective. If trusted relationships are really the most important driver of QIC success, then it may be that different QICs require different components and processes to generate this trust. For example, QICs in which individuals already know each other from previous work (e.g., QICs focused on a specific clinical discipline) may value social events or team building less than QICs in which teams are meeting for the first time. If QICs do, in fact, vary in their requirements, then future research may benefit from a taxonomic approach to QIC success where different types of QICs are identified by the specific components and processes they need to generate trust, promote open sharing, and reinforce a commitment to learning.

At the hospital-level, we treated the collaborative process not as a single intervention, but as a method for stimulating local quality improvement efforts.[11] To do this, we first imagined how the five CFIR domains might interact within a hospital performing quality improvement work and then identified domains that might be influenced by QIC participation. It is worth noting that the original CFIR model does not lay out specific connections among its five domains,[205] although other researchers have suggested possible relationships.[206] Figure 4.3 depicts our model of a hospital performing quality improvement in the absence of a QIC (CFIR domains are shaded, associated factors are not). Starting with the right column, the act of identifying a *quality problem* leads to the design and implementation of a *quality intervention*. Interventions typically target *clinical processes*, but they can occasionally target *clinical structure*, which, in turn, affects clinical processes.[12] Our model suggests that the *implementation process*

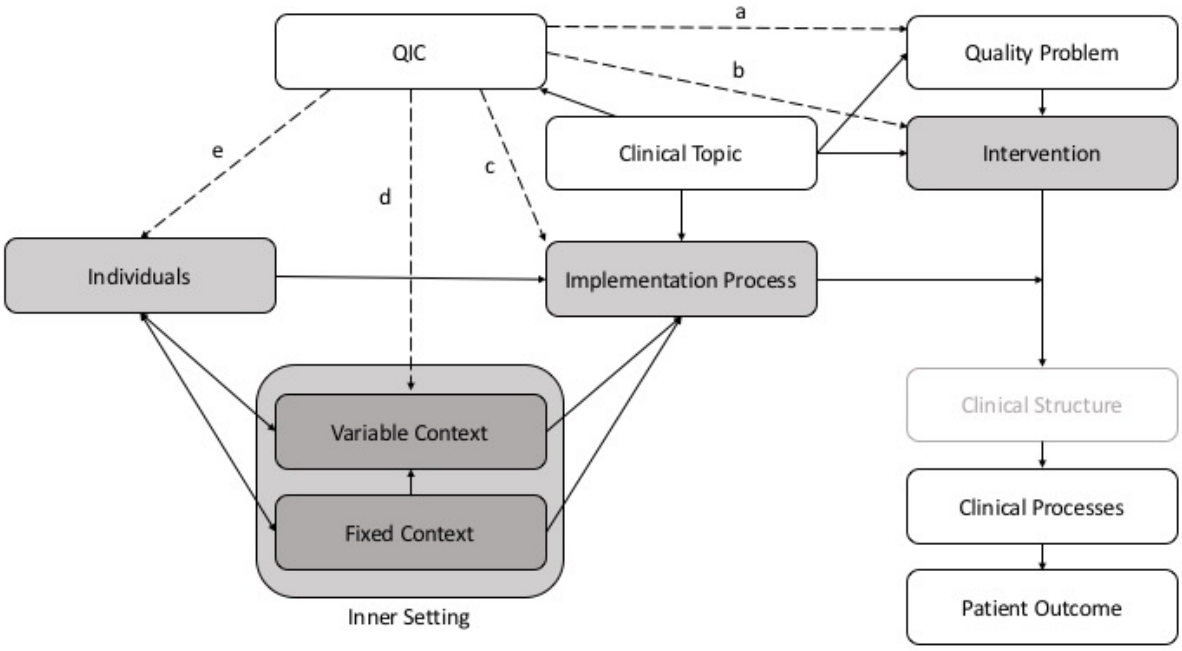
moderates the influence of the intervention on its target since the same intervention implemented in different ways can result in different results.[47,207] Although CFIR identifies the implementation process as a separate domain, we believe that it is highly influenced by both the *individuals* tasked with implementing the intervention and by the *inner setting* of the organization itself.[68,208,209] These two factors can, at the same time, influence one another: new individuals, especially senior leaders, can affect organizational culture and organizational culture can, over time, affect how individuals perform their roles.[206] We added the *clinical topic* to our model based on our interviewees' suggestion that the choice of topic may affect program success. Specifically, the existence of "what knowledge" on the topic may influence the understanding of the quality problem and the choice of intervention while the existence of "how knowledge" affects a hospital's ability to implement a solution.

Figure 4.3: A conceptual arrangement of CFIR domains



Next, we added the QIC to our baseline model and connected it to our hospital-level domains based on the benefits identified by our interviewees (Figure 4.4). This process led to five specific pathways in which QICs influence participant hospitals (labeled "a-e" in the figure). First, QICs aid in the identification and understanding of quality problems by allowing hospitals to *learn by comparison* (e.g., process benchmarking, outcome benchmarking) and by providing *technical support* (e.g., reviewing evidence-based practices, discussions with content experts). Second, QICs influence the choice of quality intervention by creating a platform for the sharing of best practices (*transfer of learning*), by connecting hospitals to experts (*technical support*), and by crowd-sourcing tests of change (*geometry of testing*).

Figure 4.4: QICs' influence on hospital-level CFIR domains



Third, QICs affect the implementation process by exposing hospitals to different implementation strategies (*outsider perspective*), by communicating lessons from

previous successes and failures (*transfer of learning*), and by providing experts to assist with common barriers or setbacks (*technical support*). Fourth, QICs affect certain aspects of the inner setting. Here, we separated aspects of an organization that are amenable to change in the short term (*variable context*: e.g., team composition, communication among microsystems) from aspects that could only be changed under sustained pressure over the longer term (*fixed context*: e.g., salary structures, organizational culture). Based on our interviewees' sense that organizational culture typically affects QICs rather than the other way around, we envisioned QICs as influencing variable context, but not fixed context. Specifically, QICs improve *accountability*, challenge fixed beliefs (*outsider perspective*), and increase clinician engagement by providing them with the opportunity to influence care for the entire region (*authority*). Finally, QICs affect individuals by increasing their technical and experiential knowledge (*technical support* and *transfer of learning*) and by making the entire quality improvement process more manageable and more enjoyable (*social support and motivation*). This final pathway of influence may affect hospital quality in and of itself since other research has suggested that teams made up of more knowledgeable and more satisfied individuals are actually more likely to produce significant improvements.[187,210]

3. **Intra-QIC variation is as much an issue of "innovation-system fit" as it is of innate "innovation attributes."** By this, we mean that hospital success in a QIC is only partially determined by the collaborative process's ability to produce meaningful outputs; a hazier but potentially more important determinant of success is how well a particular QIC's outputs meet the needs of its member hospitals and how well a hospital is able to

translate the benefits of the QIC experience into changes to its structure or clinical processes.

To explore this idea further, we first assumed that individual hospitals possess relative strengths and weaknesses in each CFIR domain and that these relative strengths and weaknesses can influence the likelihood that a hospital's quality improvement programs result in better patient outcomes.[190] For example, a hospital with competent individuals, a supportive inner setting, and a well-chosen intervention may still fail to improve because they lack the skills and attitudes necessary to successfully carry out the implementation process.

Next, we assumed that the choice of clinical topic may itself determine the relative importance of each domain. Some topics, like reducing wait times at outpatient clinics, may require a particularly robust inner setting with strong connections among clinical departments and visible support from senior leadership. Other topics, like ensuring that antibiotics are re-dosed during long operations, may require less institutional buy-in so long as the individuals involved are seen as strong and credible leaders and their positions allow them to control the necessary care processes.

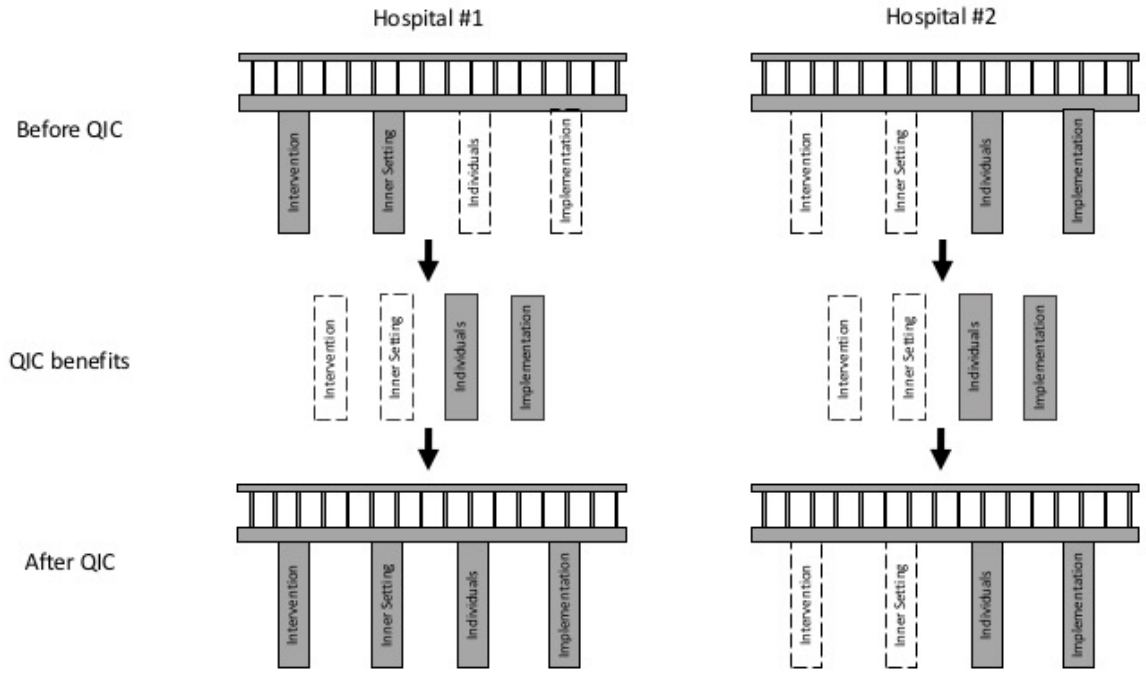
Putting these two assumptions together allows us to at least partially explain intra-QIC variation in success. Greenhalgh and colleagues suggest that innovations are more likely to be assimilated if they fit within the an organization's "existing values, norms, strategies, goals, skill mix, supporting technologies, and ways of working."[211(p608)] We

take this idea of *innovation-system fit* one step further by suggesting that hospitals benefit from QICs when their outputs supplement hospital-level deficiencies in one of the CFIR domains. For example, a QIC that predominantly influences the skills and attitudes of individuals (e.g., one that provides a high level of training in quality improvement techniques) may result in a high levels of improvement for a hospital that struggles in that area, but lower levels of improvement for a hospital whose major weakness is its disconnected and non-supportive inner setting. Since our model also suggests that the choice of topic influences the relative importance of these domains, the same two hospitals participating in the same QIC might achieve different levels of success if the collaborative had decided to pursue clinic wait times (a more inner-setting-dependent improvement topic) instead of surgical site infections (a more individuals-dependent improvement topic).

We analogize this arrangement to pillars along the base of a bridge. A bridge missing two of its four primary pillars might still be able to support the weight of a few light crossings, but not a consistent or heavy load just as a hospital that is proficient in two of the four CFIR domains might be able to perform a few straightforward quality improvement projects, but not anything requiring large-scale care redesign. If a QIC can help to fill in a hospital's missing support system, then it can make a hospital stronger and more capable of tackling complex quality efforts (Figure 4.5). If, however, a QIC only offers a hospital the same set of skills that it already has (or if hospital is unable to translate QIC benefits into changes in the organization), then it is still left without a complete support system and its performance is less likely to improve. Since each QIC is

made up of a variety of hospitals with different strengths and weaknesses, but, as a whole, produces the same set of benefits for all hospitals in the group, some will have their needs met and will demonstrate significant improvements while others will not and their performance will remain the same. It is, therefore, the variation in hospitals' needs and in hospitals' ability to capitalize on QIC benefits that produces what we observe as intra-QIC variation in success.

Figure 4.5: Intra-QIC variation stems from matching QIC benefits to hospital deficiencies



The figure depicts two hospitals participating in the same QIC. Hospital #1 knows which intervention to implement and has a supportive internal setting, but lacks the individuals or the implementation process to carry out the work. Hospital #2 has the individuals and experience with implementation, but is implementing the wrong intervention and does not have a support internal setting. Both enter the same QIC, which excels in training individuals and teaching teams how to implement quality interventions, but does not provide hospitals with the most appropriate clinical intervention or with tools to improve their internal setting. In this situation, hospital #1's quality needs have been met by the QIC and it now has a full complement of quality tools, at least for the particular topic the QIC has chosen to work on. Despite starting with similarly shaky footing and participating in the same QIC, hospital #2's quality needs endure and it is no better off at the end of the collaborative process than at the beginning. At the QIC level, one hospital has improved and another has not, the very definition of intra-QIC variation in success.

Limitations

Our study has several important limitations. First, given our focus on establishing the general storyline of QICs rather than producing an exhaustive literature review, we may have overlooked articles that provide an alternative understanding of QIC function. This is particularly true given the multidisciplinary nature of the literature on QICs. We limited our search to medical and health services journals, but relevant articles may have been published by researchers in other fields (e.g., business, organizational management). Second, our group of interviewees, while diverse and well-qualified, may not represent the full range experiences with QICs. For example, we tended to select experts that had participated in surgical QICs because of our own familiarity with the field. Moreover, while our interviewees reflect a variety of disciplines, physicians may be overrepresented and non-physician team members (e.g., nurses and administrators) may be underrepresented in our sample. Third, the term "QIC" can refer to a variety of collaborative learning systems. We focused on more formal arrangements that approximated the original Breakthrough Series model and our results may not apply to other types of multi-institutional quality efforts. Finally, our model was designed as a framework for future research on QICs and should be considered exploratory at best. Little objective data exists to support many of our hypotheses, although we hope this will change as a result of our research. Still, given the lack of any previous conceptual model on the topic, we believe that our work will benefit researchers, hospital leaders, and quality experts with an interest in QICs.

Implications for Future QICs

Our findings suggest three important modifications to our current understanding of QICs. First, trying to identify specific QIC components or processes that occur in high-performing

QICs and attempting to insert them into lower-performing QICs is unlikely to improve performance. Such a strategy is no more than low-road transfer: it falsely assumes both that these components and processes actually result in higher performance and that the same components and processes are required for success in every QIC. Instead, components and processes appear to be proxies for more important determinants of success, namely trusted relationships, open sharing, and a genuine commitment to learning. More importantly, because each QIC is made up of different hospitals with different practices, individuals, and quality needs, each QIC may need a different structure or different group activities to build trust, promote sharing, and encourage learning. This means that in measuring and predicting QIC success we should stop focusing on more tangible aspects of the group (e.g., how many members it has) and start finding ways to assess more ethereal but more important group characteristics (e.g., how those members interact).

Second, more emphasis needs to be placed on hospitals' readiness for change at the outset of the QIC process. Although our model suggests two potential sources of failure--failure of the QIC as a whole to produce meaningful outputs and failure of individual hospitals to capitalize on those outputs--interviewees felt that the second, hospital-level source was by far the most common and that hospitals' inability to capitalize on the QIC process was based primarily on poor organizational readiness for change. There is currently no widely-accepted method for assessing organizational readiness for change, although several tools have been proposed.[212-214] Our findings suggest two ways in which these tools could be applied to hospitals participating in QICs in order to improve our understanding of QIC success. First, organizational readiness for change could be measured cross-sectionally to determine if hospitals with higher levels of readiness at the beginning of the QIC process are more likely to demonstrate improvements in

patient outcomes. Such a finding might suggest that QIC admission should be limited to hospitals that meet certain readiness criteria or at least that hospitals with low levels of readiness might require additional support during the process. Second, readiness for change could be measured longitudinally to determine if participation in a QIC actually builds organizational capacity. If so, QICs might be marketed to hospitals both as short-term tools for gaining assistance with quality improvement projects and as longer-term investments in organizational development. If, however, QICs do not build capacity, then hospitals with low readiness for change scores might reconsider devoting their resources to QIC participation and instead focus on activities that more directly affect change commitment and change efficacy.

Finally, QIC organizers may benefit from explicitly placing hospitals into QICs based on their strengths and weaknesses at implementing quality improvement interventions. Unlike a pharmaceutical trial where the unit of analysis is human physiology, which varies little from patient to patient (especially in highly controlled populations), QICs target hospitals and hospital processes, which vary widely.[44] This heterogeneity makes it difficult to separate the efficacy of an intervention like the QIC (i.e., how it would perform under ideal and controlled circumstances) from its effectiveness (i.e., how it performs under real world conditions).

Currently, high levels of intra-QIC variation in capacity make designing a single collaborative process that will benefit all hospitals particularly challenging. Again, our findings suggest two ways in which the matching of QICs with hospital needs could be improved. First, QICs could restrict entry or intentionally place certain hospitals in certain QICs based on their inherent quality needs. This process already occurs on some level with self-selection of hospitals into QICs based on common interests. A more deliberate and potentially more effective matching procedure would require designing and validating an instrument for measuring hospitals' quality

needs; our model suggests that CFIR domains might serve as a basis for this assessment. Second, QICs could formally assess hospitals' quality needs at the outset and then tailor their activities to the needs of their member hospitals, even designing specific activities for specific hospitals in the group. Although more difficult, an individualized approach to QIC design might help certain hospitals make the most of their QIC experience, which could, in turn, reduce intra-QIC variation and improve the effectiveness of the QIC approach as a whole.

Conclusions

QICs have been used as a method of improving hospital quality since the late 1980s, yet many aspects of QICs--particularly how they work and why they vary in their effectiveness--remain poorly understood. Our findings suggest that QICs offer hospitals a series of unique benefits, but that these benefits are often insufficient to meet hospitals' specific quality needs. Future QICs may benefit from a more active and individualized approach to collaboration in which hospitals' readiness for change is formally measured and collaborative activities are chosen based on their ability to build trust, promote sharing, and encourage a commitment to learning.

Chapter 5 Injury-Specific Variables Improve Risk Adjustment and Hospital Quality Assessment in Severe Traumatic Brain Injury

Abstract

Background: Hospital benchmarking is essential to quality improvement, but its usefulness depends on the ability of statistical models to control for inter-hospital differences in patient mix. We explored whether the addition of injury-specific clinical variables to the current American College of Surgeons-Trauma Quality Improvement Program (TQIP) algorithm would improve model fit.

Study Design: We analyzed a prospective registry containing all adult patients who presented to a regional consortium of 14 trauma centers between 2010-2011 with severe traumatic brain injury (TBI). We used hierarchical logistic regression and stepwise forward selection to develop two novel risk-adjustment models. We then tested our novel models against the current TQIP model and ranked hospitals by their risk-adjusted mortality rates under each model to determine how model selection affects quality benchmarking.

Results: 734 patients met inclusion criteria. Stepwise selection resulted in two distinct models: one that added three TBI-specific variables (pupil reactivity, cerebral edema, loss of basal cisterns) to the 15 variables currently used by TQIP and another that combined the same three TBI-specific variables with a three-variable subset of TQIP (age, Injury-Severity Score, Glasgow Coma Scale score). Both novel models significantly outperformed the current TQIP model. Although hospital rankings remained largely unchanged across model configurations, both novel models identified fewer negative outliers.

Conclusion: The inclusion of injury-specific variables improves risk adjustment for patients with severe TBI. TQIP should consider replacing several of its general patient characteristics with injury-specific clinical predictors to increase efficiency, reduce the risk of overfitting, and improve the accuracy of hospital benchmarking.

Keywords: statistical models, risk adjustment, benchmarking, outcome assessment, hospital quality

Introduction

The American College of Surgeons has developed a variety of subscription-based benchmarking programs to help hospitals improve their quality. When performed correctly, benchmarking, which involves comparing risk-adjusted clinical outcomes among centers, allows hospitals to identify areas of weakness, learn from higher-performing centers, and track their own performance over time.[17] However, the value of benchmarking as a quality improvement tool is largely dependent on the accuracy and reliability of the underlying statistical models and their ability to adjust for patient differences across hospitals.[35]

Most benchmarking programs use a single statistical model with a general set of parameters (e.g., age, sex, functional status) to address the underlying risk of all patients instead of developing different models with different variables for each specific condition or operation (e.g., node status in colon cancer or intracranial pressure in traumatic brain injury [TBI]). Using a general set of variables reduces administrative costs and allows for direct comparisons across conditions, but can lead to imperfect model fit and inaccurate benchmarking.[215,216] Adding variables to these models can improve fit by controlling for important unmeasured differences between patients, but this strategy increases the cost of data collection, validation, and cleaning and puts the model at risk for overfitting, which may also lead to inaccurate benchmarking.[217] The optimal balance between general and condition-specific model specifications remains unknown.

Using data from our own regional trauma collaborative, we explored the importance of injury-specific variables to risk adjustment in severe TBI. To do this, we first developed two novel risk-adjustment models by integrating a set of TBI-specific clinical variables into the risk-adjustment algorithm used by the American College of Surgeons-Trauma Quality Improvement

Program (TQIP), the current gold standard for hospital benchmarking in trauma. We then validated the performance of our new models among a split-sample cohort of patients who sustained severe TBI in Los Angeles County and compared hospitals' quality rankings under these different models to better understand the influence of model selection on the results of quality benchmarking. In so doing, we hoped to understand both whether the addition of injury-specific variables improves risk adjustment and how these variables should be combined with more general patient characteristics in order to improve the accuracy of TQIP-based hospital benchmarking.

Methods

Data Source, Study Design, and Case Identification

The Los Angeles County Trauma Consortium was formed in 2013 as a collaboration between Los Angeles County's 14 trauma centers, the County's Emergency Medical Services Agency, and health services researchers from two local universities. Specifics of the consortium have been described elsewhere.[30,218] In Los Angeles County, over 200 separate clinical variables are routinely abstracted from the medical record by trained trauma program managers and entered into the Trauma and Emergency Management Information System (TEMIS), an electronic database maintained by the Emergency Medical Services Agency since 1984. All variable definitions in TEMIS comply with the National Trauma Data Standard and align with similar variables in the TQIP database. TEMIS includes a variety of pre-hospital, procedural, and injury-specific variables that are not currently captured by TQIP.

Prior to the formation of our consortium in 2013, all 14 member hospitals developed a prospective registry that included all patients presenting to any trauma center in Los Angeles

County with severe TBI between January 1, 2010 and December 31, 2011. These data form the basis of our current analysis. For the purposes of the registry, severe TBI was defined as: 1) blunt head trauma, 2) Glasgow Coma Scale (GCS) score ≤ 8 on arrival, and 3) abnormal intracranial findings on initial head CT. We excluded patients who died on arrival or were < 18 years old at the time of injury. Among the 14 hospitals, one is a designated children's hospital and was eliminated from our analysis based on the age restriction.

Variables

Our outcome of interest was death from any cause during the initial hospitalization for TBI ("inpatient mortality") as determined by the discharge destination variable in TEMIS. We selected 32 potential covariates for model development: 21 of the 22 variables in TQIP's current risk-adjustment algorithm and 11 novel variables based on the opinion of our clinical specialists in trauma surgery, neurosurgery, and neuro-critical care.

TQIP currently uses 22 variables in its regression model, including demographics, vital signs, mechanism of injury, Abbreviated Injury Scale scores, GCS, and binary variables for 12 separate medical comorbidities (see Table 5.1).[215] From this list, we were forced to make three changes based on data availability. First, since all patients in Los Angeles County are taken directly to a trauma center and there is no re-triage based on center designation (i.e., Level I vs. Level II), we ignored transfer status. Second, as we had no access to Abbreviated Injury Scale scores, we replaced this variable with the Injury Severity Score (ISS), a composite measure based on the Abbreviated Injury Scale score of the three most-severely injured body regions. Third, we used separate dummy variables for smoking and alcoholism rather than combining them as substance abuse. Age, race, systolic blood pressure (SBP), heart rate, and mechanism of

injury were treated as categorical variables based on TQIP definitions; GCS and ISS were included as continuous variables.

To this list, we added 11 TBI-specific variables collected during the study period: pupil reactivity, international normalized ratio (INR), and 9 separate intracranial findings on initial head CT (subdural hematoma, subarachnoid hemorrhage, intraparenchymal contusion, cerebral edema, intracranial hematoma, mass effect, loss of basal cisterns, epidural hematoma, and loss of grey/white differential). Pupil reactivity was defined as present or absent based on initial neuro- or trauma surgery examination. INR values were divided into normal (≤1.4) and elevated (>1.4) based on initial measurements. Head CT findings were coded independently as present or absent based on the official radiology report.

Model development

We performed non-random, split-sample validation based on the recommendations of the TRIPOD initiative.[219] This technique involves dividing our sample by year of presentation (i.e., 2010 vs. 2011), using one subsample for model development (2010), and saving the other for model validation (2011). Splitting our sample in this way reduces the risk of developing a model that fits well in a particular patient population, but generates systematically biased predictions in other patient samples, a statistical phenomenon known as overfitting.

We used a form of stepwise forward selection to build two separate multivariable mixed effects logistic regression models for inpatient mortality. Our first model (the *TQIP and TBI* model, shortened hereafter as the "*And* model") was developed by first including all TQIP variables *a priori* and then attempting to add TBI variables to the model, one at a time, until the additional variables no longer increased the model's predictive power. For our second model (the

TQIP or TBI model, shortened as the "*Or* model"), we started with an empty set of covariates and then attempted to add either TQIP or TBI variables until the model achieved its maximum predictive power.

To determine which variable to add at each step in model development, we created a series of test models by adding each remaining covariate separately to the previous model specification and then selecting the variable that produced the model with the lowest Akaike Information Criteria (AIC) value. We then performed a likelihood ratio test comparing this new model with one additional variable to the previous model specification. If this test showed a significant improvement to model fit, then this variable was added to the initial model configuration and the process was repeated, adding variables one at a time until there was no longer an improvement to model fit (i.e., the likelihood ratio test was not significant). At this point, we considered our model configuration to be final. Due to the concern for multiple testing, we chose a more conservative value ($p<0.01$) as our level of significance for all likelihood ratio tests. All models accounted for the clustering of patients within hospitals using hospital-level random effects.

We performed several sensitivity analyses to check the robustness of our model selection process. First, we used a penalized LASSO regression model starting with the entire set of covariates to simultaneously shrink regression coefficients and eliminate non-significant predictors.[220] Second, we compared regression coefficients from our hierarchical logistic regression model to coefficients calculated under both an ordinary logistic regression model and our LASSO model. As neither approach yielded different results, we present data from our initial AIC-based stepwise selection model (Appendices 5.1 and 5.2).

Model validation

We validated both final model configurations (*And* and *Or*) by applying the regression coefficients derived from our development sample to patients in our validation sample. Using this approach, we were able to calculate and compare the area under the receiver operating curve (AUROC) for our models in the two separate sample populations. We used bootstrapping with 1,000 repetitions to generate bias-corrected, 95% confidence intervals (CIs) for each AUROC calculation. Validation was considered successful if the CIs from the development sample and the validation sample overlapped. We explored the risk of overfitting using two approaches: 1) by comparing the underling distribution of variables between the two samples; and 2) by performing a separate multivariable mixed effects logistic regression on our validation sample using the same model configuration and comparing the resulting regression coefficients to those derived from the development sample. All calculations and comparisons between the development and validation samples were performed separately for the *And* and the *Or* models.

Comparison of models

After developing our novel models, we compared model fit for four possible model configurations using our entire sample population (i.e., 2010 and 2011 data): 1) a model that included only the full set of TQIP variables (*Standard TQIP*); 2) the *And* model; 3) the *Or* model; and 4) a model that included all potential covariates (i.e., the full set of TQIP and TBI variables; *All variables*). We used both AIC and AUROC values as markers of model fit and bootstrapped bias-corrected, 95% CIs for each parameter; models whose CIs did not cross were considered to have significantly different performance metrics.

To determine whether the inclusion of TBI-specific variables would affect hospital benchmarking, we compared hospital rankings under each model configuration using random effects coefficients as our marker of hospitals' risk- and reliability-adjusted performance.[221,222] Standard errors for these coefficients were calculated using the delta method and were used to determine if each hospital's risk-adjusted performance fell significantly above or below the group average. Statistical analyses were performed using STATA/SE 14.0 (StataCorp: College Station, TX) except for our LASSO model, which was performed using SAS 9.4 (SAS Institute, Inc.: Cary, NC). This study was approved by the Office of Human Research Protection Program at the University of California, Los Angeles.

Results

During the study period, 753 adult patients sustained severe TBI; 19 patients (2.5%) died upon arrival and were excluded from further analysis. Of the 734 patients who survived to admission, the mean age was 46.3 years, 24.1% were female, and 40.3% were Hispanic (Table 1). The median ISS was 26 (interquartile range [IQR] 21-35) and the median GCS was 3 (IQR 3-6). Fall was the most common mechanism of injury (32.0%) followed by auto-versus-pedestrian (28.8%). Nearly 2/3 of patients had at least one reactive pupil on arrival (65.7%). Overall, 296 patients (40.3%) died during their inpatient hospitalization.

Our development sample consisted of the 355 (48.4%) patients injured between January 1 and December 31, 2010; 379 (51.6%) patients were injured between January 1 and December 31, 2011 and formed the basis of our validation sample. Only one variable (diabetes) occurred at a different rate between the two time periods (32.8% of patients in 2010 vs. 67.2% in 2011, p=0.009; see Appendix 5.3). Six variables were excluded from the variable selection procedure

Table 5.1: Distribution of Potential Covariates by Patient Outcome

	Full Sample (n=734)	Died (n=296)	Did Not Die (n=438)	p-value (Died vs. Did Not Die)[a]
TQIP Variables				
Age (years)				<0.0001
18-25	155 (21.1)	39 (13.2)	116 (26.5)	
26-35	124 (16.9)	43 (14.5)	81 (18.5)	
36-55	219 (29.8)	72 (24.3)	147 (33.6)	
56-65	95 (12.9)	49 (16.6)	46 (10.5)	
66-75	56 (7.6)	34 (11.5)	22 (5.0)	
76-85	52 (7.1)	35 (11.8)	17 (3.9)	
86+	33 (4.5)	24 (8.1)	9 (2.1)	
Female	177 (24.1)	97 (32.8)	80 (18.3)	<0.0001
Race				0.170
Hispanic	296 (40.3)	112 (37.8)	184 (42)	
White	254 (34.6)	107 (36.1)	147 (33.6)	
Black	88 (12)	28 (9.5)	60 (13.7)	
Asian	63 (8.6)	35 (11.8)	28 (6.4)	
Other	27 (3.7)	12 (4.1)	15 (3.4)	
Unknown	6 (0.8)	2 (0.7)	4 (0.9)	
Mechanism of Injury				<0.0001
Fall	235 (32.0)	116 (39.2)	119 (27.2)	
Auto vs. Pedestrian	211 (28.7)	95 (32.1)	116 (26.5)	
Assault	98 (13.4)	33 (11.1)	65 (14.8)	
Motor Vehicle Collision	88 (12)	19 (6.4)	69 (15.8)	
Motorcycle Collision	59 (8.0)	18 (6.1)	41 (9.4)	
Other	43 (5.9)	15 (5.1)	28 (6.4)	
Injury Severity Score, mean (SD)	29.1 (12.5)	34.5 (12.0)	25.4 (11.4)	<0.0001
Glasgow Coma Scale, mean (SD)	4.4 (1.9)	3.7 (1.4)	4.9 (2.0)	<0.0001
Pre-hospital Cardiac Arrest[b,c]	8 (1.1)	8 (2.7)	0 (0.0)	-
Heart Rate (beats per minute)[d]				0.505
0-89	270 (36.8)	111 (37.5)	159 (36.3)	
90-119	281 (38.3)	116 (39.2)	165 (37.7)	
120+	183 (24.9)	69 (23.3)	114 (26.0)	
Systolic Blood Pressure (mm Hg)[d]				0.183
0-99	74 (10.1)	43 (14.5)	31 (7.1)	
100-149	367 (50.0)	120 (40.5)	247 (56.4)	
150-199	234 (31.9)	97 (32.8)	137 (31.3)	
200+	59 (8.0)	36 (12.2)	23 (5.3)	
TQIP Comorbidities				
Hypertension	142 (19.3)	70 (23.6)	72 (16.4)	0.020
Alcoholism	94 (12.8)	24 (8.1)	70 (16.0)	0.002
Diabetes	67 (9.1)	37 (12.5)	30 (6.8)	0.013
Current Smoker	40 (5.4)	10 (3.4)	30 (6.8)	0.057
Obesity	15 (2.0)	7 (2.4)	8 (1.8)	0.602
Respiratory Disease	14 (1.9)	6 (2.0)	8 (1.8)	0.839

Heart Disease	11 (1.5)	6 (2.0)	5 (1.1)	0.355
Bleeding Disorder[c]	9 (1.2)	6 (2.0)	3 (0.7)	0.082
Functionally Dependent[c]	8 (1.1)	4 (1.4)	4 (0.9)	0.454
Stroke[c]	5 (0.7)	3 (1.0)	2 (0.5)	0.386
Cancer[c]	4 (0.5)	3 (1.0)	1 (0.2)	0.200
Liver Disease[c]	3 (0.4)	2 (0.7)	1 (0.2)	0.382
TBI-Specific Variables				
At Least One Reactive Pupil	482 (65.7)	123 (41.6)	359 (82.0)	<0.0001
Elevated INR (>1.4)[d]	82 (11.2)	60 (20.3)	22 (5.0)	<0.0001
Intracranial Findings on Head CT[e]				
Subdural Hematoma	483 (65.8)	225 (76.0)	258 (58.9)	<0.0001
Subarachnoid Hemorrhage	442 (60.2)	190 (64.2)	252 (57.5)	0.090
Intraparenchymal Contusion	256 (34.9)	86 (29.1)	170 (38.8)	0.005
Cerebral Edema	197 (26.8)	136 (45.9)	61 (13.9)	<0.0001
Intracranial Hematoma	169 (23.0)	71 (24.0)	98 (22.4)	0.632
Mass Effect	152 (20.7)	89 (30.1)	63 (14.4)	<0.0001
Loss of Basal Cisterns	136 (18.5)	104 (35.1)	32 (7.3)	<0.0001
Epidural Hematoma	74 (10.1)	16 (5.4)	58 (13.2)	0.001
Loss of Grey/White Differential	58 (7.9)	50 (16.9)	8 (1.8)	<0.0001

ICP, intracranial pressure; INR, international normalized ratio; SD, standard deviation; TBI, traumatic brain injury; TQIP, Trauma Quality Improvement Program
[a]P-values based on hierarchical mixed effects regression with patients clustered within treating hospitals
[b]Defined by TQIP criteria as heart rate less than 50 bpm on presentation
[c]Not included in variable selection procedure
[d]Includes imputed values
[e]Presence of intracranial findings based on radiology report from initial head CT

because they either had a prevalence of zero (cancer and liver disease) or because they perfectly predicted the outcome in at least one of the two time periods (cardiac arrest, bleeding disorder, functionally dependent, and stroke; see Appendix 5.4). A correlation matrix with all potential covariates confirmed a lack of multicollinearity, which reduced the risk that similar variables would be systematically removed during variable selection (Appendix 5.5).

Forward selection led to the addition of three TBI-specific variables to the standard TQIP model, creating an *And* model with 18 variables: the 15 remaining TQIP variables plus cerebral edema, pupil reactivity, and loss of basal cisterns (Table 5.2A). The final *Or* model contained six variables: pupil reactivity, age, ISS, cerebral edema, GCS, and loss of basal cisterns (Table 5.2B). All three TBI-specific variables from the *And* model were also contained within the *Or*

Table 5.2: Stepwise Construction of the And and Or Models

Table 5.2A: Stepwise Construction of the *And* Model

Step	Variable with lowest AIC	AIC[a]	Deviance[b]	p-value[c]	Result
1	Cerebral edema on head CT	313.42	34.47	<0.0001	Add
2	At least one reactive pupil (Yes/No)	294.19	21.22	<0.0001	Add
3	Loss of basal cisterns on head CT	287.19	7.71[d]	0.0055[d]	Add
4	Elevated INR (>1.4)	286.08	3.11	0.0776	Stop

Table 5.2B: Stepwise Construction of the *Or* Model

Step	Variable with lowest AIC	AIC[a]	Deviance[b]	p-value[c]	Result
1	At least one reactive pupil (Yes/No)	420.19	68.75	<0.0001	Add
2	Age	379.36	50.84	<0.0001	Add
3	Injury Severity Score	336.92	46.44	<0.0001	Add
4	Cerebral edema on head CT	304.17	34.74	<0.0001	Add
5	Glasgow Coma Scale score	295.35	10.82	0.0010	Add
6	Loss of basal cisterns on head CT	289.06	8.29	0.0040	Add
7	Diabetes	284.70	6.37	0.0116	Stop

AIC, Akaike information criteria; CT, computed tomography; INR, international normalized ratio
[a]AIC of baseline model with only TQIP variables was 345.89. AIC of baseline model without any covariates was 484.94.
[b]Deviance calculated as -2*(log likelihood of smaller model - log likelihood of larger model)
[c]Based on likelihood ratio chi2 test
[d]Larger model was unable to estimate a random intercept, generating a situation where the Δdf between the models is zero. The P-value for this step was, therefore, calculated by comparing two ordinary logistic regression models with the same model configuration.

model. Both final models had similar AIC values (*And* model 286.08 vs. *Or* model 289.06), suggesting that the remaining TQIP variables contributed minimally to model fit.

Table 5.3 presents the results of our split-sample model validation. Although the 95% CIs for the two samples overlapped under the *Or* model, suggesting successful validation, the *And* model had a significantly larger AUROC in the development sample than in the validation sample, suggesting overfitting. A comparison of regression coefficients between the two samples found three variables with significant associations in one sample, but not the other: gender, diabetes, and smoking. Attempts to improve fit by removing these variables from the model increased the concordance of AUROC values between the samples, but not enough to achieve validation (see Appendix 5.6).

Table 5.3: Validation of the And and Or Models

	AUROC (95% CI[a])	
	Development sample (2010 data)	Validation sample (2011 data)
TQIP and TBI model	0.942 (0.917, 0.962)	0.847 (0.801, 0.881)
TQIP or TBI model	0.915 (0.885, 0.942)	0.882 (0.846, 0.913)

AUROC, area under the receiver operator curve; CI, confidence interval
[a]95% confidence intervals bootstrapped and bias-corrected with 1,000 repetitions

Table 5.4: Fit Statistics for All Four Model Specifications

Model	Number of variables	AIC[a,b]	AUROC[b]
Standard TQIP	15	727.40 (707.36, 777.05)	0.862 (0.836, 0.889)
TQIP and TBI	18	597.13 (582.45, 654.16)	0.917 (0.897, 0.935)
TQIP or TBI	6	603.55 (610.60, 619.57)	0.900 (0.876, 0.920)
All variables	26	585.70 (581.51, 620.49)	0.925 (0.908, 0.944)

AIC, Akaike information criterion; AUROC, area under receiver operator curve; TBI, traumatic brain injury; TQIP, Trauma Quality Improvement Program
[a]For AIC, lowest represents best fit
[b]95% confidence intervals bootstrapped and bias-corrected with 1,000 repetitions

All four models performed reasonably well in our full sample, with AUROCs ranging from 0.862 to 0.925 (Table 5.4; see Appendix 5.7 for full regression coefficients from all four model specifications). As would be expected, models with more variables generally had larger AUROC estimates, except for the *Standard TQIP* model, which produced both the lowest AUROC and the highest AIC statistic. The remaining three models yielded overlapping values for both statistics, suggesting no statistical difference in model fit.

The implications for hospital benchmarking are illustrated in Table 5.5 and Figure 5.1. Risk adjustment reclassified between seven (*Or* model) and nine hospitals (*Standard TQIP* and *All variables*) to different quality terciles compared to the unadjusted data. One hospital, in

Table 5.5: Comparison of Unadjusted and Risk-adjusted Mortality Rates for Severe Traumatic Brain Injury at 13 Los Angeles County Trauma Centers, 2010-2011

Center[a]	Sample size	Unadjusted Mortality Rate, % (95% CI[b])	Risk-adjusted Mortality Rate, % (95% CI[c])			
			Standard TQIP Model	*And* Model	*Or* Model	*All Variables* Model
A	30	20.0 (4.8, 35.2)	27.8 (17.3, 41.6)	32.3 (22.3, 44.3)	33.1 (24.2, 43.5)	31.0 (20.0, 44.6)
B	58	24.1 (12.8, 35.5)	36.4 (25.2, 49.4)	41.0 (30.2, 52.6)	40.0 (30.5, 50.2)	40.9 (29.2, 53.7)
C	58	34.5 (21.9, 47.1)	38.7 (27.3, 51.5)	40.3 (30.0, 51.5)	40.9 (31.6, 50.8)	40.4 (29.0, 53.0)
D	62	35.5 (23.2, 47.7)	37.0 (26.1, 49.4)	40.7 (30.3, 52.2)	41.5 (32.1, 51.5)	42.2 (30.6, 54.8)
E	19	36.8 (13.0, 60.7)	34.8 (21.7, 50.6)	35.5 (24.4, 48.4)	36.7 (26.8, 47.8)	32.9 (21.1, 47.4)
F	72	40.3 (28.7, 51.9)	34.4 (24.4, 46.0)	36.9 (27.3, 47.7)	38.0 (29.2, 47.7)	36.7 (26.3, 48.5)
G	183	42.6 (35.4, 49.9)	52.4 (43.7, 60.9)	46.3 (37.8, 55.0)	46.6 (38.7, 54.6)	47.2 (38.0, 56.7)
H	46	45.7 (30.7, 60.6)	37.9 (26.4, 50.9)	39.4 (29.0, 50.9)	41.0 (31.4, 51.3)	39.0 (27.6, 51.8)
I	37	45.9 (29.1, 62.8)	52.2 (37.7, 66.4)	44.4 (32.5, 56.8)	40.3 (30.4, 51.0)	44.8 (31.4, 59.1)
J	52	46.2 (32.1, 60.2)	36.4 (24.9, 49.7)	37.1 (26.8, 48.7)	38.7 (29.3, 49.0)	37.1 (25.7, 50.2)
K	45	48.9 (33.7, 64.1)	47.9 (34.9, 61.2)	46.2 (34.7, 58.1)	43.9 (33.8, 54.4)	49.4 (36.1, 62.6)
L	52	50.0 (36.0, 64.1)	45.5 (33.1, 58.6)	42.1 (31.3, 53.6)	41.5 (32.0, 51.7)	40.7 (28.9, 53.7)
M	20	50.0 (26.0, 74.0)	45.8 (30.7, 61.8)	43.1 (31.0, 56.0)	43.0 (32.3, 54.3)	43.7 (30.0, 58.4)

[a]Codes are intentionally arbitrary to protect hospitals' identities. However, the order reflects hospitals' unadjusted mortality rate from low to high
[b]Confidence intervals based on the standard error for the mean under a normal distribution
[c]Confidence intervals based on the standard error of the random intercept estimated from our hierarchical model

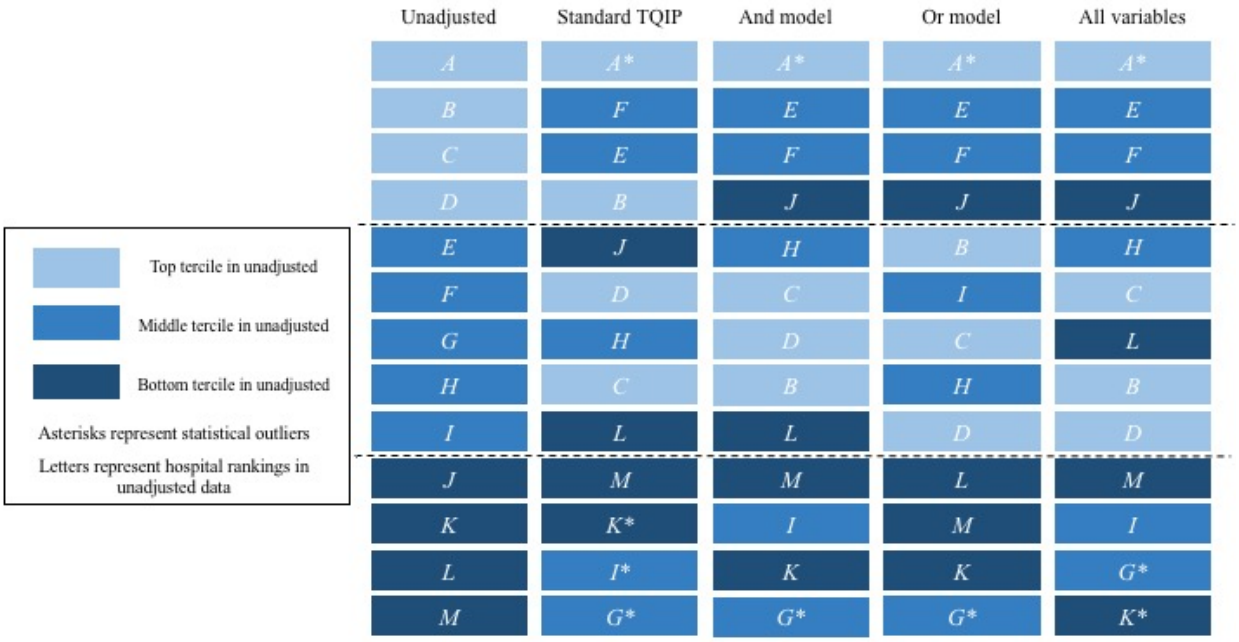
particular, moved from the lowest to the middle (*Standard TQIP*) or highest quality tercile (all remaining models) after risk adjustment. Models that controlled for TBI-specific variables demonstrated a relatively similar pattern of hospital rankings to the *Standard TQIP* model with several hospitals moving across one tercile (e.g., middle to low or high to middle), but none moving from the lowest to the highest tercile or visa versa. However, while the *Standard TQIP* model identified four statistical outliers (one positive, three negative), both the *And* and the *Or* model identified only two (one positive, one negative). The model with all 26 potential variables also identified three outliers (one positive, two negative; see Appendix 5.8 for caterpillar plots).

Discussion

Hospital benchmarking is an essential tool for quality improvement, but its usefulness depends largely on the ability of statistical models to control for inter-hospital differences in patient or case mix. Unlike traditional risk-adjustment techniques, which have focused on a large but general set of trauma-related variables, we developed two higher performing and more parsimonious model specifications by replacing general patient characteristics with injury-specific clinical predictors. Not only did these models demonstrate better fit characteristics than the standard TQIP model, both resulted in different hospital rankings for risk-adjusted mortality, even within our small group of 14 regional trauma centers.

Our study is not the first to attempt to improve risk adjustment in trauma by comparing multiple potential model specifications, although it is the first, to our knowledge, to focus on the issue of injury-specific variables. Haider and colleagues[223] used a complete sample of patients from the National Trauma Data Bank over two years (2009-2010) and a similar stepwise variable selection approach to develop a novel predictive model for mortality. Although their focus was

Figure 5.1: Comparison of Hospital Rankings for Mortality in Severe Traumatic Brain Injury by Choice of Risk-adjustment Model



on trauma patients in general rather than patients with severe TBI, there are remarkable similarities between their final model specification and our *Or* model, with age, GCS, and ISS appearing in both. In fact, the differences in model specification--Haider and colleagues' model also included hypotension, pulse, and ventilator use while ours included pupil reactivity and two head CT findings--likely reflect both variable availability (we did not have access to ventilator use) and differences in the underlying patient population (head CT findings may be less important to survival in penetrating chest trauma) rather than variations in the model building process.

At least two other studies have attempted to limit the burden of data collection by reducing the number of variables in established risk-adjustment models, although neither focused specifically on trauma. Osborne and colleagues[224] developed a seven-variable version of the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP) for patients undergoing five vascular surgery procedures while Dimick and colleagues[225]

developed both a five-variable and a two-variable version of ACS-NSQIP for patients undergoing five general surgery procedures. Both studies demonstrated similar fit statistics for their more limited models and concluded that adequate risk adjustment can be performed using smaller, more directed sets of patient characteristics. Our study seems to confirm these findings since our most limited model, the *Or* model, performed significantly better than the standard TQIP model and no worse than a model that included all possible covariates.

As we hypothesized, injury-specific variables, including clinical and radiologic findings thought to be important markers of injury severity in TBI, were among the most significant predictors of inpatient mortality in our sample. In fact, pupil reactivity had the highest predictive power of any single potential covariate and the addition of pupil reactivity and two head CT findings (cerebral edema and loss of basal cisterns) to the standard TQIP configuration significantly improved model fit. Moreover, the relatively small difference in fit between the *And* and *Or* models--roughly three AIC points--suggests that the remaining TQIP variables (i.e., every variable except for age, ISS, and GCS) contribute little to the model's predictive capacity and could safely be dropped from future risk-adjustment efforts. Doing so would almost certainly increase both the administrative and computational efficiency of TQIP by minimizing the cost of data collection, maximizing statistical power, and reducing the risk of overfitting among smaller sub-samples.

Our inability to validate the *And* model in our 2011 data, despite multiple configuration changes, suggests the presence of inflated regression coefficients on variables that were statistically overrepresented in the 2010 data, a common symptom of overfitting. Our events per variable ratio in developing the *And* model (calculated as the number of events per independent variable in the regression model) was 8.3 (149 deaths divided by 18 variables), below the typical

rule-of-thumb value of 10 that is needed to avoid overfitting.[217] Therefore, replacing more general model variables with injury-specific predictors might not only improve the accuracy of risk adjustment, but also allow for the development of more robust statistical models for hospital performance among particular subsets of patients, such as those with severe TBI.

In that vein, our finding that slightly different risk-adjustment models can result in significantly different hospital rankings raises the question of how to best measure hospital quality. Although the overall pattern of quality terciles remained largely the same across all four model configurations, both novel models identified two fewer negative outliers, suggesting that these hospitals may actually be no worse than average after adequately controlling for TBI-specific markers of injury severity. In a related study, Shafi and colleagues[226] found no difference in hospitals' performance for blunt versus penetrating injuries and, in turn, suggested that hospital quality need not be measured separately for different types of traumatic injury. However, other studies[16,227] have cautioned that observed differences in hospital performance across validated metrics may reflect the innately multifocal nature of hospital quality, such that no single metric can adequately capture quality across all of its domains. If this is, in fact, the case, then developing models specific to subsets of patients, such as those with severe TBI, may identify clinical areas within a hospital that would benefit from focused improvement efforts and may help to move the field away from the idea that hospital quality is binary (i.e., that a hospital is either good or bad). ACS-NSQIP has begun to explore procedure-specific modeling to improve the granularity of the information it provides to its member hospitals[228]; our results suggest that TQIP may benefit from a similar approach in which model configurations vary to match our knowledge of underlying clinical processes.

Our study has several important limitations. First, our sample included only the most severely-injured patients in Los Angeles County, and, as such, our results may not generalize to other regions, to less severely-injured patients, or to patients with other types of traumatic injury. Second, we had no information on the rates of care withdrawal, and, therefore, could not control for the ways in which different withdrawal practices among centers might have affected center-specific mortality rates. Third, stepwise selection does not allow for the evaluation of every possible combination of variables and there may exist other model specifications that perform better than our final configurations.[229] However, the robustness of our final specifications to multiple methods of variable selection suggests that we have identified and captured the most important predictors in each of our models. Fourth, any variable selection procedure risks including variables with a high degree of association but no plausible causal connection to the outcome. While this risk is somewhat lower in our study since all of our potential covariates were derived from clinical predictors already thought to be important to trauma patients, we may still have included variables that merely proxy for other unmeasured effects (e.g., patients with signs of cerebral edema may have simply experienced a longer delay in getting to the CT scanner). However, since our intent was to improve risk adjustment rather than to elucidate predictors of mortality in TBI, including variables with a spurious connection to our outcome would still allow for appropriate control of inter-hospital differences in patient or case mix so long as they do not replace other independent predictors of mortality. Given that our model specification with all potential covariates performed no better than either our *And* or *Or* models, we find this possibility unlikely.

These limitations notwithstanding, our findings have important implications for quality improvement in trauma. Historically, hospital benchmarking program have been initiated by

national specialty organizations like the American College of Surgeons, which benefit from a wide reach and substantial administrative resources, but may lack the flexibility to rapidly test and implement novel quality improvement strategies. More recently, however, regional consortia of hospitals, such as the Los Angeles County Trauma Consortium and the Michigan Trauma Quality Improvement Program (MTQIP)[230], have begun to share data in the hope of improving the clinical care of trauma patients. Regional collaboratives represent a unique opportunity to explore local practices in depth, to provide social support and motivation to colleagues working on similar quality problems, and to test new methods of data collection and analysis. The influence of these regional programs on national quality improvement efforts can be significant; in fact, the entire TQIP program began out of applying ACS-NSQIP principles to trauma patients at a single academic medical center in Michigan that is now the coordinating center for the new MTQIP program.[231] If regional collaboratives continue to use their smaller, more adaptive, and more collegial arrangements to serve as laboratories for quality improvement techniques, then they may continue to play a significant role in improving the way we think about and implement surgical quality improvement at the national level as well.

Conclusions

Hospital benchmarking is an essential quality improvement tool for understanding local performance, but remains only as reliable as the underlying statistical models that drive its results. We used injury-specific variables available through a regional quality collaborative to improve current risk-adjustment practices, making them both more accurate and more efficient for patients with severe TBI. If models themselves can be tailored specifically to subsets of the trauma population, then we may be able to generate more accurate assessments of hospital

quality and help hospitals focus their quality improvement efforts on the specific clinical areas in which they currently underperform.

Chapter 6 Compliance with Evidence-Based Guidelines and Interhospital Variation in Mortality for Patients with Severe Traumatic Brain Injury

Abstract

Importance: Compliance with evidence-based guidelines in traumatic brain injury (TBI) has been proposed as a marker of quality. However, the association between hospital-level compliance rates and risk-adjusted clinical outcomes in TBI remains poorly understood.

Objective: To examine whether hospital-level compliance with Brain Trauma Foundation (BTF) guidelines for intracranial pressure (ICP) monitoring and craniotomy is associated with risk-adjusted mortality in severe TBI.

Design, Setting, Participants: All adult patients (n=734) presenting to a regional consortium of 14 hospitals between 2009-2010 with severe TBI (blunt head trauma, Glasgow Coma Scale [GCS] < 9, abnormal intracranial findings on head computed tomography [CT]). We used hierarchical mixed effects models to assess the relationship between hospital-level compliance with BTF guidelines and mortality after adjusting for patient-level demographics, severity of trauma (e.g. mechanism of injury, injury severity score [ISS]), and TBI-specific variables (e.g. cranial nerve reflexes, head CT findings).

Main Outcomes and Measures: Hospital-level risk-adjusted inpatient mortality, hospital-level compliance with BTF guidelines for ICP monitoring and craniotomy

Results: Unadjusted mortality rates varied by site from 20.0-50.0%; risk-adjusted rates varied from 24.3-56.7%. Overall, only 46.1% (338/734) of patients with an appropriate indication underwent ICP monitor placement and only 45.6% (134/335) underwent craniotomy. Hospital-level compliance ranged from 9.6-65.2% for ICP monitoring and 6.7-76.2% for craniotomy. We found widespread variation in compliance across hospitals, but no association between hospital-level compliance rates and risk-adjusted patient outcomes.

Conclusions and Relevance: Hospital-level compliance with evidence-based guidelines has minimal association with risk-adjusted outcomes in TBI. Our results suggest caution should be taken before using compliance with these measures as independent quality metrics. Given the complexity of TBI care, outcome-based metrics, including functional recovery, may be more accurate than current process measures at determining hospital quality.

Keywords: traumatic brain injury, guidelines, compliance, quality metrics, mortality

Introduction

Traumatic brain injury (TBI) remains a substantial source of morbidity and mortality in the United States, accounting for nearly 1/3 of all injury-related deaths.[29,232,233] Following reports of widespread variation in care,[234] the Brain Trauma Foundation (BTF) published the first set of clinical guidelines for the treatment of TBI in 1995. In their most recent form, BTF guidelines include management strategies, treatment thresholds, and indications for the use of invasive procedures, namely intracranial pressure (ICP) monitoring--placing a small pressure monitor inside the skull--and craniotomy--the removal of a portion of the skull to allow for brain swelling.[235-237] Beyond receiving more streamlined clinical care, individuals who are treated based on BTF guidelines appear to have better clinical outcomes, including lower risk-adjusted mortality and higher rates of functional recovery.[238-240] As a result, compliance with these guidelines has been proposed as a marker of hospital quality, and hospitals throughout the country are attempting to improve TBI outcomes by increasing their levels of guideline-compliant care.[241]

However, despite substantial face validity, there is growing evidence that guideline compliance alone is an inaccurate and inadequate measure of hospital quality.[242,243] Multiple studies have demonstrated a loose relationship between compliance with Medicare's Hospital Compare measures (e.g. giving aspirin within 24 hours of an acute myocardial infarction) and inpatient mortality.[244,245] More specific to surgery, hospitals that score well on peri-operative safety measures do not appear to have lower rates of SSI,[246] venous thromboembolism,[247] or mortality.[248-250] To date, there remains no definitive evaluation of the connection between guideline compliance and hospital quality in severe TBI.

Our study had two specific aims: 1) to document levels of compliance with BTF guidelines for ICP monitoring and craniotomy within a large regional trauma system, and 2) to examine the association between hospital-level compliance with these guidelines and risk-adjusted mortality. In so doing, we sought to determine whether compliance with BTF guidelines represents a viable marker of hospital quality in severe TBI.

Methods

Study Design and Data Sources

The Los Angeles (LA) County Trauma Consortium was formed in 2013 as a collaboration between LA County's 14 trauma centers, the County's Emergency Medical Services Agency, and health services researchers from two local universities. A description of the Consortium and its research objectives has been published previously.[30] Briefly, we developed a prospective registry of all patients sustaining severe TBI in LA County over a two-year period (2009-2010). Inclusion was based on 3 criteria: (1) blunt head trauma, (2) Glasgow Coma Scale (GCS) < 9 on arrival, and (3) abnormal intracranial findings on initial head CT. These criteria were chosen based on the definition of severe TBI in BTF guidelines. We excluded patients who died on arrival or were < 18 years old at the time of injury. Among the 14 hospitals in the Consortium, one is a designated children's hospital and was effectively eliminated from our analysis based on the age restriction.

Trained trauma program managers at each center prospectively identified patients meeting inclusion criteria and abstracted relevant patient-level data into our protected, electronic registry. Hospital characteristics were obtained from 2 sources. Basic structural characteristics were taken from the American Hospital Association website (available at

http://www.ahadataviewer.com). Trauma program managers were also surveyed regarding care practices at their institution during the study period, specifically whether there was a designated neurocritical care unit, a general or neurosurgery residency program, or a clinical protocol for the treatment of severe TBI.

Patient-level Variables

Our primary outcome at the patient level was death from any cause during the initial hospitalization for TBI ("inpatient mortality"). ICP monitor placement within the first 72 hours after arrival was recorded as an independent field in our clinical registry. Craniotomy during the first 72 hours was determined based on *International Classification of Diseases, Ninth Edition* (ICD-9) procedure codes (Appendix 6.1).

Covariates were selected to parallel risk-adjustment techniques used by the Trauma Quality Improvement Program (TQIP), the largest clinical registry of traumatically-injured patients in the U.S. For TBI patients, TQIP includes 23 variables in its regression models: demographics, vital signs, mechanism of injury, Injury Severity Score (ISS), GCS, and 14 medical comorbidities. Three variables (dialysis, concurrent steroid use, and cardiac arrest) were recorded but not used in statistical analyses due to their low prevalence and high collinearity.

In addition to the 20 TQIP variables, we included 11 TBI-specific variables captured by our registry: pupil reactivity, international normalized ratio (INR), and 9 separate intracranial findings on initial head CT (Table 6.1). Pupil reactivity was defined as present or absent based on the initial neuro- or trauma surgery consultation. INR values were divided into normal ($\leq$1.4) and elevated (>1.4) based on initial measurements in the emergency department. All 9 head CT findings were coded independently as present or absent based on the official radiology report.

Hospital-level Quality Metrics

We determined guideline compliance for each hospital by dividing the number of patients receiving a particular therapy (i.e. ICP monitoring or craniotomy) by the number of patients with an indication for that therapy based on BTF guidelines. BTF guidelines currently recommend ICP monitoring for all TBI patients with a GCS < 9;[235] as this includes all patients in our sample, we divided the number of patients undergoing ICP monitoring at a given hospital by that hospital's sample size. BTF guidelines also recommend craniotomy for patients with a GCS < 9 in the setting of epidural hematoma, subdural hematoma with signs of midline shift or mass effect, or intraparenchymal contusion with signs of mass effect.[237] We identified patients at each hospital meeting these criteria, and then calculated rate of patients undergoing craniotomy in this group.

Risk-Adjusted Mortality

We developed a hierarchical mixed-effects logistic regression model to predict inpatient mortality after controlling for all 31 patient-level covariates (20 TQIP + 11 TBI-specific variables) and hospital-specific random effects. We predicted risk-adjusted mortality rates for each hospital after controlling for patient-mix, and then used empiric Bayesian techniques to adjust each estimate for its reliability.[221,222,251] This approach, also referred to as "shrinkage" adjustment, filters out statistical noise due to the small sample size of certain clusters and reweights values toward the overall sample mean based on each estimate's reliability.[228]

Statistical Analysis

Bivariate tables were generated to compare patients who died during their hospitalization to those who survived to discharge. For unadjusted data, we used chi-squared tests of independence for categorical variables and either two-sample t tests or Mann-Whitney U tests for continuous variables, depending on the distribution. We also ranked hospitals by their compliance with BTF guidelines, and then divided hospitals into terciles based on these rankings. This process was done separately for ICP monitoring and craniotomy such that the distribution of hospitals within terciles could differ for each quality metric.

We evaluated the association between hospital-level compliance rates and risk-adjusted mortality in two manners. First, we plotted hospitals' compliance rates against their risk-adjusted mortality rates and calculated Spearman rank correlations. This was done separately for each quality metric. Second, we added hospitals' guideline compliance in terciles to our hierarchical regression model and predicted the risk-adjusted mortality rate for each tercile. This was also done separately for each quality metric using only patients who were eligible for that measure. We then compared unadjusted mortality rates across terciles using the Kruskal-Wallis test and risk-adjusted mortality rates using Wald tests with standard errors generated via the delta method. All p-values were two-sided and levels ≤ 0.05 were considered significant.

We conducted several sensitivity analyses to test the robustness of our estimates. First, we calculated the reliability of our registry data by comparing our independent data field for ICP monitoring to ICD-9 procedure codes (Appendix 6.2). As this demonstrated a high level of consistency (Cohen's $\kappa = 0.852$), we used our independent data field. Second, 43 patients (5.7%) were found to have missing data in 2 variables: INR (36 patients) and heart rate (7). Multiple analyses demonstrated no informative pattern to the missing data. In order to use all available data, we imputed missing values via a multivariate normal regression technique after controlling

for the full set of covariates, including ICP monitor placement, craniotomy, and mortality. As an additional test of our findings, we performed our analyses using only complete cases. This did not affect our results and we report regression results for the full sample with imputed values for patients with missing data. Third, due to the concern than Bayesian techniques might inappropriately pull smaller high-performing or low-performing centers toward the group mean,[252] we compared regression coefficients and observed-to-expected mortality ratios, another measure of hospital performance, between our hierarchical logistic regression and an ordinary logistic regression that did not account for clustering (Appendices 6.3 and 6.4). As these analyses showed no significant differences in regression coefficients, hospital rankings, or hospital outlier status, we report the results for our hierarchical logistic regression model. Fourth, we restricted our analyses to patients with an isolated head injury (Abbreviated Injury Score < 2 in all body regions except head/neck); this did not affect our findings and we report data for the entire sample. Finally, we performed our regression analyses using hospital compliance as a continuous variable. As this also did not affect our results, we report terciles for interpretability. Statistical analyses were performed using STATA/IC 13.0 (StataCorp: College Station, TX). This study was approved by the Office of Human Research Protection Program at the University of California, Los Angeles.

Results

During the study period, 753 adult patients sustained severe TBI; 19 patients (2.5%) died upon arrival and were excluded from further analysis. Of the 734 patients who survived to admission, the mean age was 46.3 years, 24.1% were female, and 40.3% were Hispanic (Table 4.1.1). The median ISS was 26 (interquartile range [IQR] 21-35) and the median GCS was 3

Table 6.1: Characteristics of Severe TBI Patients Admitted to Los Angeles County Trauma Centers, 2009-2010

	Full Sample (n=734)	Died (n=296)	Survived (n=438)	p-value
Demographics				
Age (years), mean (SD)	46.3 (20.7)	53.7 (21.9)	41.3 (18.2)	<0.001
Female	177 (24.1)	97 (32.8)	80 (18.3)	<0.001
Race				0.045
Hispanic	296 (40.3)	112 (37.8)	184 (42.0)	
White	254 (34.6)	107 (36.2)	147 (33.6)	
Black	88 (12.0)	28 (9.5)	60 (13.7)	
Asian	54 (8.6)	32 (11.8)	22 (6.4)	
Other	27 (3.7)	12 (4.1)	15 (3.4)	
Unknown	6 (0.8)	2 (0.7)	4 (0.9)	
Injury and Physiologic Response				
Mechanism of Injury				<0.001
Fall	235 (32.0)	116 (39.2)	119 (27.2)	
Auto vs. Pedestrian	211 (28.8)	95 (32.1)	116 (26.5)	
Assault	98 (13.4)	33 (11.2)	65 (14.8)	
Motor Vehicle Collision	88 (12.0)	19 (6.4)	69 (15.8)	
Motorcycle Collision	59 (8.0)	18 (6.1)	41 (9.4)	
Other	43 (5.9)	15 (5.1)	28 (6.4)	
Injury Severity Score, median (IQR)	26 (21-35)	30 (25-50)	25 (17-33)	<0.001[§]
Glasgow Coma Scale, median (IQR)	3 (3-6)	3 (3-4)	4 (3-7)	<0.001[§]
Systolic Blood Pressure (mm Hg), mean (SD) [†]	143.2 (38.2)	143.6 (46.8)	142.9 (31.1)	0.802
Heart Rate (beats per minute), mean (SD) [†]	99.0 (30.1)	96.6 (31.8)	100.7 (28.9)	0.007
Comorbidities				
Hypertension	142 (19.4)	70 (23.7)	72 (16.4)	0.015
Alcoholism	94 (12.8)	24 (8.1)	70 (16.0)	0.002
Diabetes	67 (9.1)	37 (12.5)	30 (6.9)	0.009
Current Smoker	40 (5.5)	10 (3.4)	30 (6.9)	0.042
Obesity	15 (2.0)	7 (2.4)	8 (1.8)	0.613
Respiratory Disease	14 (1.9)	6 (2.0)	8 (1.8)	0.846
Heart Disease	11 (1.5)	6 (2.0)	5 (1.1)	0.333
Bleeding Disorder	9 (1.2)	6 (2.0)	3 (0.7)	0.105
Functionally Dependent	8 (1.1)	4 (1.4)	4 (0.9)	0.575
Stroke	5 (0.7)	3 (1.0)	2 (0.5)	0.368
Cancer	4 (0.5)	3 (1.0)	1 (0.2)	0.156
Liver Disease	3 (0.4)	2 (0.7)	1 (0.2)	0.351
TBI-Specific				
At Least One Reactive Pupil	482 (65.7)	123 (41.6)	359 (82.0)	<0.001
Elevated INR (>1.4)[†]	83 (11.3)	61 (20.6)	22 (5.0)	<0.001

Intracranial Findings on Head CT[+]				
Subdural Hematoma	483 (65.8)	225 (76.0)	258 (58.9)	<0.001
Subarachnoid Hemorrhage	442 (60.2)	190 (64.2)	252 (57.5)	0.071
Intraparenchymal Contusion	256 (34.9)	86 (29.1)	170 (38.8)	0.007
Cerebral Edema	197 (26.8)	136 (46.0)	61 (13.9)	<0.001
Intracranial Hematoma	169 (23.0)	71 (24.0)	98 (22.4)	0.611
Mass Effect	152 (20.7)	89 (30.1)	63 (14.4)	<0.001
Loss of Basal Cisterns	136 (18.5)	104 (35.1)	32 (7.3)	<0.001
Epidural Hematoma	74 (10.1)	16 (5.4)	58 (13.2)	0.001
Loss of Grey/White Differential	58 (7.9)	50 (16.9)	8 (1.8)	<0.001
Outcome and Quality Metrics				
Inpatient Mortality	296 (40.3)	-	-	
ICP Monitor Placement	338 (46.1)	110 (37.2)	228 (52.1)	<0.001
Craniotomy[^]	134 (40.0)	56 (30.8)	78 (50.9)	<0.001

All numbers represent percentages unless otherwise stated.
^Among eligible patients: epidural hematoma, subdural hematoma with CT signs of midline shift or mass effect, and intraparenchymal contusion with CT signs of mass effect.
†Includes imputed values.
§Based on Mann Whitney U test.
+Presence of intracranial findings based on radiology report from initial head CT.
ICP, intracranial pressure; INR, international normalized ratio; SD, standard deviation; TBI, traumatic brain injury

(IQR 3-6). Fall was the most common mechanism of injury (32.0%) followed by auto-versus-pedestrian (28.8%). Nearly 2/3 of patients had at least one reactive pupil on arrival (65.7%). Overall, 296 patients (40.3%) died during their inpatient hospitalization.

The majority of trauma centers were not-for-profit (69%) and maintained an academic affiliation (39% major, 31% minor; Table 2). Less than half had a neurocritical care unit, a general or neurosurgery resident program, or a protocol for treating severe TBI during the study period. Unadjusted mortality rates ranged from 20.0 to 50.0% by hospital (median 42.6, IQR 35.5-46.2); risk-adjusted rates ranged from 24.3 to 56.7% (median 41.1, IQR 36.4-47.8).

Only 338 patients (46.1%) underwent ICP monitor placement despite all patients meeting BTF criteria. Rates of compliance with ICP monitoring ranged by center from 9.6 to 65.2% (median 47.4, IQR 40.0-51.7); the lowest tercile of hospitals had a median compliance rate of 38.7% compared to 54.8% in the highest tercile (Table 6.2). Among our sample, 335 (45.6%)

Table 6.2: Characteristics of Hospitals in the L.A. County Trauma System, 2009-2010

	N (%)
General structural factors[#]	
Ownership	
Government	4 (30.8)
Not-for-profit	9 (69.2)
Teaching status	
Major	5 (38.5)
Minor	4 (30.8)
None	4 (30.8)
Total beds, median (IQR)	371 (318 - 466)
Admissions per year, median (IQR)	20,398 (17,117 - 23,508)
Structural factors related to TBI [‡]	
Neurocritical Care Unit	5 (38.5)
General Surgery Residency Program	6 (46.2)
Neurosurgery Residency Program	4 (30.8)
Protocol for the Treatment of TBI	6 (46.2)
TBI quality metrics	
Use of ICP Monitoring, group median (range)	
Lowest Tercile (n=223)[*], %	38.7 (9.6 - 43.2)
Middle Tercile (n=304), %	48.5 (46.7 - 51.4)
Highest Tercile (n=207), %	54.8 (51.7 - 65.2)
Use of Craniotomy, group median (range)[^]	
Lowest Tercile (n=330) , %	9.5 (6.7 - 37.2)
Middle Tercile (n=212), %	42.6 (41.7 - 45.7)
Highest Tercile (n=192), %	61.7 (50 - 76.2)

#Data from Hospital Profiles on American Hospital Association Data Viewer website (available at http://www.ahadataviewer.com).
‡Data from survey of trauma program managers.
^Among eligible patients: epidural hematoma, subdural hematoma with CT signs of midline shift or mass effect, and intraparenchymal contusion with CT signs of mass effect.
*Sample sizes reflect patients within terciles. Because the division was performed at the hospital level, the numbers of patients are not equal across terciles. At the hospital level, both groupings had 5 hospitals in the lowest, 4 hospitals in the middle, and 4 hospitals in the highest tercile.
ICP, intracranial pressure; IQR, interquartile range; TBI, traumatic brain injury

patients had an indication for craniotomy and 134 of these (40.0%) underwent the procedure. Rates of compliance with craniotomy also varied by center from 6.7 to 76.2% (median 41.7, IQR 33.3-50.0); the lowest tercile of hospitals had a median compliance rate of only 9.5% compared to 61.7% in the highest tercile.

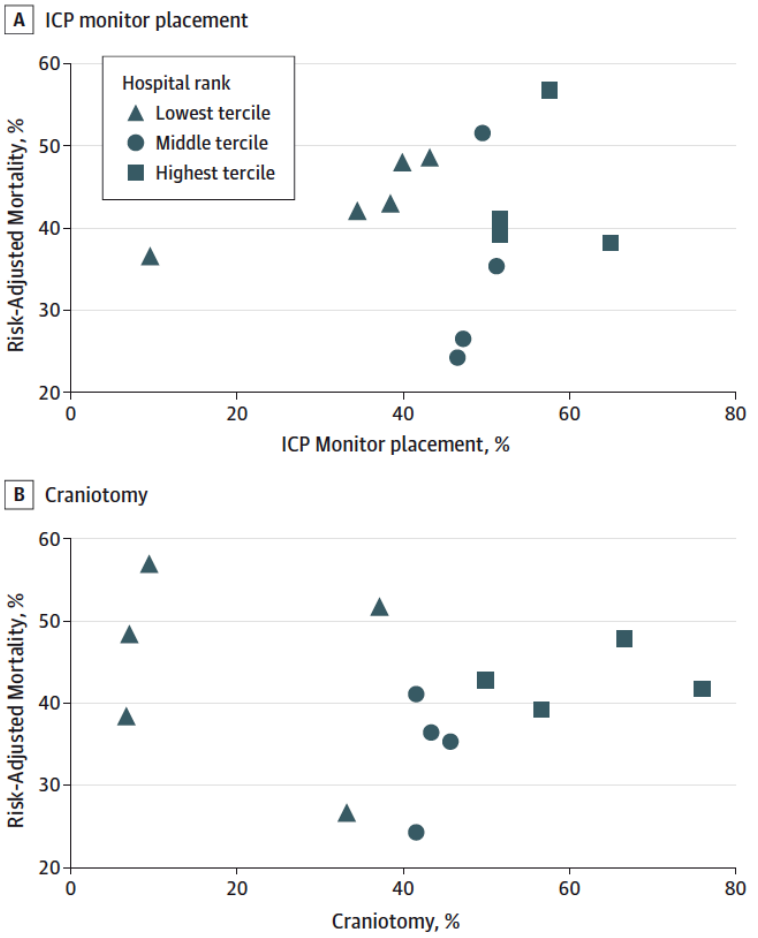
Table 6.3: Association Between Hospital-Level Compliance with Quality Metrics and Inpatient Mortality

	Unadjusted mortality rate, %	p-value*	Risk-adjusted mortality rate,° %	p-value#
Use of ICP Monitoring		0.292		0.230
Lowest Tercile	44.4		41.8	
Middle Tercile	39.5		33.8	
Highest Tercile	37.2		42.0	
Use of Craniotomy^		0.347		0.446
Lowest Tercile	58.1		55.8	
Middle Tercile	51.2		47.1	
Highest Tercile	49.4		56.0	

^Among eligible patients: epidural hematoma, subdural hematoma with CT signs of midline shift or mass effect, and intraparenchymal contusion with CT signs of mass effect.
*Results from Kruskal-Wallis test comparing rates across terciles.
°Based on hierarchical logistic regression model predicting inpatient mortality after controlling for all 31 patient-level variables and hospital-level random effects (n=734 for ICP monitoring, n=335 for craniotomy).
#Results from Wald test comparing rates across terciles based on standard errors derived using the delta method.

Figure 6.1 presents each hospital according to its rates of guideline compliance and risk-adjusted mortality. There was no correlation between compliance rates and mortality for either ICP monitoring (Spearman $\rho = 0.030$, $p = 0.922$; Figure 6.1 [A]) or craniotomy (Spearman $\rho = -0.066$, $p = 0.830$; Figure 6.1 [B]). Similarly, there was no association between guideline compliance and either unadjusted or risk-adjusted mortality based on our hierarchical regression model (Table 6.3). For both quality metrics, there appeared to be a trend toward lower mortality in hospitals with higher compliance based on unadjusted data, however, this disappeared after risk-adjustment. The middle tercile of compliance had the lowest risk-adjusted mortality rates for both ICP monitoring (33.8%) and craniotomy (47.1%).

Figure 6.1: Risk-adjusted Mortality Rates by Hospital Performance on Quality Metrics



A, Hospitals according to their rates of intracranial pressure (ICP) monitor placement and risk-adjusted mortality. B, Hospitals according to their rates of craniotomy and risk-adjusted mortality.

Discussion

In response to widespread variation in care, BTF established clinical practice guidelines for TBI in the hope of promoting high-quality care. Despite their efforts, we found no relationship between hospitals' compliance with two BTF guidelines and risk-adjusted mortality within a large regional trauma system. Our results suggest caution should be taken before using compliance with these measures as independent quality metrics.

While ours is the first study to explore differences in hospitals' use of craniotomy, other studies have investigated the association between hospital compliance with ICP monitoring and risk-adjusted mortality in severe TBI.[253,254] Using data from TQIP, Alali and colleagues reported improved survival for patients treated at hospitals in the highest compared to the lowest quartile of ICP monitor use after controlling for patient- and hospital-level covariates.[253] There are several possible explanations for our differing results. First, Alali and colleagues excluded patients with either "non-survivable" head injuries or significant non-head injuries. Since we included all patients with severe TBI, our sample was more-severely injured by comparison, as is evident by the difference in unadjusted mortality rates (35.5% in TQIP vs. 40.3% in our sample). As a result, certain invasive treatments may have been less effective in our population, making inter-hospital differences in their use a less significant predictor of patient outcomes. Second, TQIP includes mostly large, academic hospitals, which tend to be more similar to one another than to other hospitals within a given geographic region. Of our 13 adult trauma centers, only 3 currently participate in TQIP (all Level 1, academic centers) and fewer than half have access to many of the hallmarks of academic medical centers, including specialized ICUs or residency programs. Given the variety of hospitals involved, differences in guideline compliance may simply play a smaller role in determining hospital quality for regional trauma systems--especially considering that differences in ICP monitoring explained less than 10% of the inter-hospital variation in mortality in TQIP.[253]

Our results also appear to conflict with multiple patient-level analyses--including one from our Consortium--that have consistently demonstrated a survival benefit for ICP monitoring in severe TBI.[255,256] Again, there are at least two possible explanations for this discrepancy: (1) differences in *who* receives ICP monitoring, and (2) how effectively information from invasive

monitors is incorporated into clinical management. First, even among the subset of patients with severe TBI, there is considerable heterogeneity such that not all patients benefit equally from invasive monitoring.[257] Therefore, two hospitals with similar rates of ICP monitoring might effectively be providing very different levels of care--and achieving very different patient outcomes--if the subset of patients being monitored differs significantly between the two institutions.

Second, by recording only *if* an ICP monitor was placed, guidelines may insufficiently capture critical aspects in the management of severe TBI, including how clinicians react to changes in ICP or use related medical therapies (e.g. hypertonic saline) to treat intracranial hypertension. Much of the literature surrounding ICP monitoring suggests that clinical outcomes depend more on the timely management of intracranial hypertension than the use of particular medications or procedures.[258-260] This idea is typified by the only RCT of ICP monitoring, which showed no difference in mortality between monitored and non-monitored patients, but took advantage of trained neurologists and neurosurgeons at the bedside acting as "functional ICP monitors."[261] Therefore, if hospitals that place more ICP monitors do not also use the additional information to adapt their treatment of intracranial hypertension, then the rate of ICP monitor placement may be uncorrelated (or even inversely correlated) with risk-adjusted mortality.

Our study has several important limitations. First, our sample included the most severely-injured patients in LA County, and, as such, our results may not generalize to other regions or to less severely-injured patients. Second, due to the delay between data collection and analysis, it is possible that care practices have changed over time. However, as neither the guidelines nor the members of the trauma system changed during this period, we believe our results remain applicable to the current state of trauma care in LA County. Third, our sample size was relatively

small at the hospital level, which limited our ability to compare hospitals directly and to control for hospital-level structural variables, such as teaching status or patient volume. As other studies have suggested these factors may impact quality,[262,263] our results may be confounded if differences in structural characteristics drive variation in both compliance and mortality. However, as BTF guidelines apply equally to patients regardless of their treating hospital, we believe that controlling for certain hospital characteristics is not only irrelevant, but may bias our results by explaining away meaningful differences in patient outcome. Additional work using a larger hospital-level sample is needed to better understand the relationships between structural variables, inpatient mortality, and guideline compliance. Fourth, we had no information on the rates of care withdrawal, and, therefore, could not control for the ways in which different withdrawal practices between centers might have affected center-specific mortality rates. Finally, we were unable to characterize patients' functional status at the time of discharge, and it is unclear whether differences in the use ICP monitoring and craniotomy are associated with differences in functional recovery.

These limitations notwithstanding, we believe our results have important implications for quality assessment in severe TBI. The debate between process and outcome has continued since the earliest publications on quality measurement.[264] Process measures--including guideline compliance--offer the ease of measurement, sensitivity to change, and real-time feedback demanded by many proponents of quality improvement. Their downside, however, rests in the difficulty determining which processes actually lead to better care.[265] Our results suggest that current BTF guidelines fail to improve hospital quality both by measuring the wrong processes (i.e. *if* a monitor is placed rather than *how* a monitor is used in management) and by doing so in

the wrong group of patients (all patients with severe TBI rather than only those who stand to benefit from invasive monitoring).

However, recent experiences from several large, multi-institutional quality improvement programs suggest that even tracking the "perfect" process measure in the most appropriate subset of patients may be insufficient to actually improve patient outcomes.[9,246,266] These findings have 3 implications for quality measurement, especially for conditions as complex as TBI. First, multiple, related process measures may be needed for each clinical outcome (e.g. ICP monitor placement *plus* thresholds for treatment *plus* cerebral perfusion pressure goals).[173,267] Second, process measurement must be combined with clinically-important outcome metrics, such as mortality and functional recovery, to prevent hospitals from gaming the system by improving performance on one metric at the expense of others. Finally, it is not enough to simply measure performance and track progress over time.[268] Instead, hospitals must build systems that use these data to influence physician behavior and improve patient care.

Conclusions

Despite improvements in care, mortality from TBI remains both common and variable from hospital to hospital. Our results demonstrate no association between hospitals' compliance with two BTF guidelines and risk-adjusted mortality, suggesting that neither measure should be used as an independent marker of hospital quality.

Chapter 7 A Time and Place for QICs

Improving the quality of healthcare remains central to the future of American medicine. Despite multiple calls to action over the past decade and a half,[180-182,269] the U.S. healthcare system continues to rank last among 11 developed nations in one performance metric[270] and 37th out of 191 countries in a separate analysis by the World Health Organization.[271] When one combines these discouraging statistics with the fact that many high-profile improvement efforts have been unable to reproduce their success[9] and that clinical registries--the current in-vogue strategy--appear to be better at tracking change than at supporting it,[26,27] it is hard not to conclude that our current approach to quality improvement is not working. The question now is: what do we need to do differently in order to reliably and sustainably improve care?

The purpose of this dissertation has been to explore and evaluate one particular strategy: the quality improvement collaborative. In theory, QICs build on the apparent weaknesses of our current improvement strategies, namely single-institution CQI and clinical registries. Unlike single-institution CQI, where hospitals are generally left on their own to design and implement improvement programs, QICs provide an external expert- and peer-based support system. In Chapter 4, I examined the differences between single-institution CQI and QICs in depth and presented eight unique benefits of the QIC approach, the majority of which centered around its ability to provide focused support at each step in the improvement process. A similar contrast can be made between QICs and clinical registries, which offer some benefits of the QIC approach (e.g., using multi-institutional, risk-adjusted data to identify quality problems), but fail to provide the same level of support during the implementation process. As Osborne and Etzioni write in their commentary on NSQIP, "participation in a quality measurement platform is…only 1 part of a comprehensive approach to quality improvement. For real quality improvement to

take place, outcomes data need to be used to guide changes in practice."[28] To date, clinical registries have largely failed to actually transform clinical practices, at least in the way that QICs were designed to do.

Yet, despite the theoretical benefits of QICs, there is no evidence to suggest that they consistently improve patients' health. In Chapter 2, I reviewed the published literature on QICs in depth and found that they appear to work in some settings, but not others. In Chapter 4, I took this idea further by trying to identify QIC- and hospital-level factors that are associated with success. With Chapters 5 and 6, I took a different approach and explored ways in which QICs might provide value, not necessarily to individual hospitals, but to the field of quality improvement as a whole. Taken together, my research suggests that, while QICs offer some unique benefits to both hospitals and to the field, they are, by no means, a panacea. In fact, QICs suffer from many of the same problems as single-institution CQI, namely that hospitals must first understand their own processes and then fully commit themselves to change in order to be successful. This is perhaps why Plsek, in his initial paper on QICs, writes that, "collaborative improvement efforts do not replace an organization's quality management efforts; rather, they depend and build on them."[11(p89)] Since many hospitals still struggle with establishing a readiness and commitment to change, I believe that QICs have been and will continue to be limited in their effectiveness, at least until this capacity can be built through other mechanisms.

Given that all three improvement strategies have unique strengths and weaknesses, I will conclude with a brief discussion on when and where each strategy might be deployed based on my own definition of quality and the various building blocks of improvement work that other quality researchers have identified.

Components of quality

In Chapter 1, I proposed a definition for quality with three central components. Yet, in evaluating QICs, the current literature has focused, almost exclusively, on whether or not QIC participation results in changes to specific process or outcome measures. While such an approach is useful for comparative effectiveness research (e.g., do QICs improve care more than clinical registries?), it may miss other ways in which QICs affect hospital quality.[272] Although it is well beyond the scope of this final chapter to propose a new method for evaluating QICs, I do want to briefly revisit my definition of quality to better understand how each improvement strategy may affect different components of quality. In so doing, I hope to better understand the true value of each improvement strategy such that hospitals can choose the most appropriate strategy based on their own quality goals (Table 7.1).

Component 1: Quality is determined, at least in part, by what we as physicians do and do not do for and to our patients. With this component, I had hoped to focus our attention on whether or not QICs actually influence hospital-level care processes, which I split into two dimensions (technical and interpersonal) based on the prevailing quality theory.[12,13] As I discussed at length in Chapter 2, QICs do appear to affect technical aspects of quality, although whether they do so more than single-institution improvement efforts remains hotly debated. In CQI, hospitals typically start by evaluating their own processes to identify gaps in care and then test different quality interventions until they find one that fits their needs.[17] In QICs, the process is similar, except that participant hospitals typically use outcome benchmarking to identify gaps in care and then attempt to implement processes associated with high-performing hospitals. Only for clinical registries is the focus on the processes of care indirect. Receiving a registry report may identify clinical areas in which a hospital underperforms (e.g., surgical site infection after

colectomy), but it cannot directly tell hospital leaders what it is about their care processes that is producing these deficiencies. Further work is needed to translate registry findings into better care, which must be done either by drilling down on one's own processes (as in single-institution CQI) or by learning how to improve care from other hospitals (as in QICs).

Table 7.1: Strategies' influence on my three components of quality

	Component 1: Influence processes of care	**Component 2:** Document and explain variation	**Component 3:** Embrace multiple perspectives
Single-institution CQI	X	Intra-hospital only	X
Clinical registries	Indirect only	Inter-hospital only	
QICs	X	X	X

As a corollary to component 1, I mentioned that, since processes can be split into technical and interpersonal, QICs could theoretically improve quality by addressing either dimension. Interestingly, I found no example of a QIC that directly addressed interpersonal processes (e.g., how physicians communicate with patients' families) and only a few that addressed downstream markers of interpersonal care, like patient experience (e.g., reducing clinic wait times[98,273]). To the best of my knowledge, none of the major clinical registries directly address interpersonal aspects of quality, although several single-institution programs have attempted to explore and improve communication between providers and patients facing difficult surgical decisions.[274,275] If QICs can, in fact, influence interpersonal quality (e.g., by leveraging close relationships between individuals in the group), then they may offer another advantage over clinical registries. If not, then hospitals working to improve non-technical aspects of quality may be forced to rely upon single-institution CQI.

Component 2: Quality varies in ways that do not reflect differences in disease severity or patient mix. With this component, I suggested that QICs could improve quality by

either documenting or helping to explain variations in care. Again, based on the limited data that do exist, it appears that QICs can and do perform this function. While several large QICs have helped to document gaps in clinical outcomes,[121,133] the real value of QICs rests in understanding the sources of this variation. For example, site visits during the NNECVDSG helped surgeons understand that differences in technical care, hospital organization, and peri-operative decision-making were contributing to inter-hospital differences in mortality.[92,195,276] Similarly, group feedback during the Michigan Keystone Project allowed organizers to change the focus of the intervention from a standard, checklist-based protocol to a collective effort to improve safety culture.[47] In both instances, the collaborative environment allowed for a deeper understanding of how hospitals differed in their processes of care and why these differences came to exist in the first place.

Neither single-institution CQI nor clinical registries provide as complete a look into the mechanisms underlying variation in care. Single-institution efforts can certainly help individual hospitals understand its own variation (e.g., different environmental services teams may be more or less efficient at turning over ORs between cases), but their findings may not generalize to other institutions or help to explain differences among hospitals. Similarly, clinical registries can help to expose inter-hospital variation in outcomes, but do little to link this variation to specific care processes and may actually conceal important sources of intra-hospital variation by linking all providers to the same hospital-level performance metric.

Component 3: There is no uniform way of defining or measuring quality; in fact, its definition often depends on the perspective of the interested party. With this final component, I had hoped to focus on the idea that what that represents high quality to one group of people might not represent high quality to another. Here, I believe QICs set themselves apart

from clinical registries and, in some ways, from single-institution CQI as well. Among the most commonly cited benefits of the QIC approach, both in the published literature and in my interviews with experts, is its flexibility to meet the needs and interests of the group. In describing regional QICs like MSQC, Campbell et al. write that, "flexibility is indeed a crucial feature of the regional collaborative. This means flexibility to make midcourse corrections and flexibility to innovate."[121(pS54)] Kwon et al. similarly write that, "regional initiatives like SCOAP offer a flexible, adaptive, and locally sensitive complement to national programs."[134(p152)]

Unlike national registries, QICs allow participants to set the agenda, choose the target, discuss which interventions should be considered, and even collect whatever variables they feel are most relevant to the group. One interviewee even went so far as to say:

> The idea that a national group would dictate the direction our group would take was not viewed favorably. We felt like coming back to them and saying, "listen, this is our state. These are our hospitals and we may have a different interest in quality improvement in this state than what it is nationally."

Multiple QICs have created their own performance metrics and used data in ways that are not available to hospitals participating in a centralized program like NSQIP or TQIP. As I described in Chapter 5, my own experience with LACTC suggests that even the best national programs may miss important data elements that can improve risk adjustment and even alter hospital benchmarking. More than simply acknowledging the interests of the participants, the flexibility to collect new data or perform new analyses can actually help hospitals identify gaps in care while the flexibility to alter evidence-based protocols can help hospitals adapt interventions to local context.

In addition to their flexibility, QICs and single-institution CQI are specifically designed to incorporate the perspectives of multiple disciplines. (Although clinical registries involve nurses as data collectors, they, like other local providers, often play a much smaller role in

deciding what actually gets measured and how those data are analyzed.) In QICs, multidisciplinary teams traditionally consist of physicians, administrators, and senior leaders,[60] but can occasionally include nurses or other hospital employees.[277] For CQI, the list can be much longer since hospitals are specifically tasked with identifying and involving every individual who directly or indirectly affects patient care.[18] Although a larger group can be more difficult to manage, having the broadest possible perspective on a quality problem may help hospitals identify root causes or specific areas in need of improvement. Therefore, as with efforts to address the interpersonal aspects of quality, if QICs fail to allow all of the necessary voices to be heard, then hospitals might benefit from turning to single-institution CQI in order to deepen their understanding of a particularly stubborn or complex quality problem.

Knowledge systems

Another way to compare the three improvement strategies is by what each offers to hospitals that are attempting to improve care. Batalden and Davidoff describe five separate knowledge systems that must be combined in order to produce a successful improvement effort: 1) scientific evidence, 2) context awareness, 3) adaptation, 4) execution, and 5) performance measurement.[278] Since all three strategies rely on performance measurement in roughly the same way (i.e., to track progress over time or to compare performance to other institutions), I will focus my comparison on the other four knowledge systems. Again, my hope is that, through this comparison, hospitals will be better able to select the most appropriate improvement strategy for their quality needs.

Scientific evidence. Both QICs and clinical registries can help to generate scientific evidence about a given clinical problem, although the national generalizability and statistical

power of clinical registries gives them a slight advantage over smaller, regional QICs (Table 7.2). In fact, many QICs intentionally pursue quicker, less statistically rigorous analyses in order to maintain the momentum needed for rapid-cycle improvement programs. [69,279] This may be why QICs appear to be particularly reliant on pre-developed content and implementation knowledge; at their core, they are methods for disseminating knowledge rather than developing it. Single-institution CQI programs, on the other hand, rarely affect the general scientific knowledge on a particular clinical issue both because they tend to favor simpler and more visual data formats to traditional regression techniques and because their data collection is designed specifically for the hospital at hand.[280]

Context awareness. As I discovered in Chapter 4, QICs' success--and the success of quality improvement programs in general--relies, in large part, on the readiness and the ability of participant hospitals to actually carry out specific changes to clinical care. A central component of this readiness is an awareness of one's own workflow: what are the patterns of care, who are the important care providers, and how do they they interact with the rest of the organization? [200] Here, QICs offer some benefit, primarily in forcing hospitals to analyze their own data, teaching them how to think about process variation, and providing a framework for comparison. However, the depth with which hospitals come to understand their own context is considerably less in QICs than in single-institution CQI, where the entire improvement process is centered around process redesign. This idea is borne out in my interviews with QIC experts, many of whom felt that context awareness was both central to hospitals' success in QICs and wholly unaffected by the collaborative process itself. In their minds, hospitals that do well in QICs have already developed a keen sense of their own organizational context, typically by attempting multiple single-institution improvement projects prior to joining the QIC. In contrast, registries provide little-to-

no direct support for hospitals in terms of understanding their own clinical practices. Although NSQIP provides access to some resources on quality improvement and puts together an annual conference for hospitals to discuss their findings, its primary model remains providing hospitals with risk-adjusted clinical outcomes and hoping each will use this data to direct its own improvement efforts.[25]

Table 7.2: Strategies' influence on Batalden and Davidoff's five separate knowledge systems

	Scientific evidence	Context awareness	Adaptation	Execution	Performance measurement
Single-institution CQI		XX	XX		X
Clinical registries	XX				XX
QICs	X	X	X	XX	XX

XX = strong influence, X = weak influence, empty = no influence

Adaptation. The difficulties that several highly publicized and well-funded improvement efforts have faced in replicating their success in new environments only highlights a central tenant of modern quality improvement theory: even the best quality intervention must be adapted to fit local context. Batalden and Davidoff describe this as "applying and adapting generalisable evidence to particular contexts."[278(p2)] Single-institution CQI generally requires little-to-no adaptation; although components may be borrowed from other quality programs, most CQI-based interventions are generated within the institution itself and are based on directly an in-depth examination of its current practices. While QICs were designed to help lower-performing hospitals adapt and implement practices from high-performing hospitals in the group, the extent to which they actually do this depends on how the collaborative is run and on how participant hospitals process the information that is generated through data sharing.[59] If hospitals are able to translate each other's experiences rather than simply copying them (i.e., if they adopt high-road

versus low-road transfer[197]), then QICs may play a large and important role in merging technical and experiential knowledge with local context. This is not the case, however, with clinical registries, which, despite also being multi-institutional groups of hospitals, do not explicitly examine differences in clinical practice or attempt to translate interventions from one setting to another. If anything, clinical registries often attempt to make all hospitals look the same (e.g., by encouraging the widespread adoption of certain evidence-based guidelines) rather than trying to determine the ideal intervention and the ideal set of clinical practices for each individual hospital.

Execution. It is QICs' influence on the actual execution or implementation of quality interventions that sets them apart from either of the other improvement strategies in my mind. As I discussed in Chapter 4, in CQI, teams may undergo some training in quality improvement techniques or may hire an outside consultant, but they are typically on their own throughout the implementation process. Hospitals that have either a particularly experienced quality staff,[205] a strong organizational culture,[281] or a firm belief that the work is in their own strategic interest[34] may be able to succeed without any additional support; the majority of hospitals, however, have more difficulty, which helps to explain the minimal impact CQI appears to have had on hospital quality more generally.[35] Clinical registries also typically provide little-to-no assistance to hospitals with respect to the actual implementation process. Shared forums, annual meetings, and online resources are often made available to hospitals, but there is not the same level of access to implementation experts or even to one's own peers in a clinical registry as there is under the QIC approach.

A combined approach?

Putting these two lines of comparison together allows for better picture of what each improvement strategy offers. For quality problems in which best practices have yet to be identified, a clinical registry allows hospitals to begin to build an evidence base and to hone in on organizations that appear, for some reason, to be outperforming others in the group. Conversely, for quality problems in which there is already a high level of scientific evidence and a small set of clinical interventions appear to be both effective and easy to implement, a single-institution CQI approach may be not only be adequate, but potentially more efficient than any strategy requiring outside parties. In fact, one interviewee said:

> If we're talking about something like implementing the surgical checklist where there is a lot of literature and it's clear what it is you're supposed to do and you've got supportive surgeons and they understand that it's not just checking boxes but implementing change based on what those checks look like, then [hospitals] can probably do it on their own.

Interestingly, he went on to say:

> And, if they don't make rapid progress, they might say, "well, this is not as easy as we thought," and join a collaborative.

This idea of starting with one approach (e.g., single-institution CQI) and then moving to another (e.g., a QIC) indicates that the three improvement strategies are by no means mutually exclusive. In fact, the different strengths and weaknesses I have identified suggest that all three strategies can and perhaps should be combined in order to have the largest impact on quality. In a combined approach, clinical registries, given their large sample size, data reliability, and experience with performance measurement, could be used as the primary source of data. However, rather than simply being used to produce semi-annual reports, these data would need to be shared with regional groups of hospitals and used as a framework for additional, process-oriented data collection. Such a design would play to each strategy's competitive advantage (i.e.,

clinical registries for outcomes benchmarking and QICs for process benchmarking), avoid redundancy, and allow local groups to control and direct improvement efforts toward their particular areas of interest.

But even this combined "registry-QIC" framework may not be enough. In fact, similar arrangements are already in the works for a number of NSQIP-based QICs, although the extent to which these groups actually receive, manipulate, and add to the pooled data they receive from NSQIP varies greatly from QIC to QIC. As I have demonstrated through this dissertation, the success of QICs depends, in large part, on individual hospitals' interest, ability, and experience with quality improvement. For that reason, I believe that encouraging single-institution CQI and other capacity-building programs--not instead of, but in conjunction with clinical registries and QICs--may be a necessary first step in overcoming many of our most persistent barriers to improvement.[282] There is still no magic bullet, but continuing to develop the QIC approach and integrating it into the larger armamentarium of improvement tools gives us the best chance to improve the quality of American healthcare for good.

Appendices

Appendix 2.1: Search strategy from Schouten and colleagues

DATABASE SEARCHED & TIME PERIOD COVERED:

PubMed, May 2012 – October 2015

LANGUAGE:

English

SEARCH STRATEGY:

(quality[tiab] AND improvement[tiab] AND collaborative[tiab]) OR (series OR project) AND breakthrough)

AND

(("Organizational Innovation"[Mesh] OR "Cooperative Behavior"[Mesh] OR "Models, Organizational"[Mesh]) AND ("Program Evaluation"[Mesh]) OR "Total Quality Management"[Mesh]) OR "Quality Assurance, Health Care"[Mesh]) AND ("Outcome and Process Assessment (Health Care)"[Mesh]) AND ("Health Services Research"[Mesh]) OR "Regional Medical Programs"[Mesh])

TOTAL: 169

Appendix 2.2: Search strategy to identify surgical QICs

DATABASE SEARCHED & TIME PERIOD COVERED:

PubMed, prior to March 2015

LANGUAGE:

English

SEARCH STRATEGY:

(quality[tiab] AND improvement[tiab] AND collaborative[tiab]) OR ((series OR project) AND breakthrough)

AND

[("Organizational Innovation"[Mesh] OR "Cooperative Behavior"[Mesh] OR "Models, Organizational"[Mesh])

OR ("Program Evaluation"[Mesh] OR "Total Quality Management"[Mesh] OR "Quality Assurance, Health Care"[Mesh])

OR ("Outcome and Process Assessment (Health Care)"[Mesh]) AND ("Health Services Research"[Mesh] OR "Regional Medical Programs"[Mesh])]

AND

[("Surgery Department, Hospital"[Mesh] OR "Operating Rooms"[Mesh] OR "Surgical Procedures, Operative/organization and administration"[Mesh] OR "Perioperative Care/organization and administration"[Mesh])

OR

(surgery[tw] OR surgeries[tw] OR surgical[tw] OR operative[tw] OR intraoperative[tw] OR preoperative[tw] OR perioperative[tw] OR "operating room"[tw] OR "operating rooms"[tw] OR "operating theatre"[tw] OR "operating theater"[tw] OR "operating theatres"[tw] OR "operating theaters"[tw])]

TOTAL: 119

Appendix 2.3: Search strategy to identify articles or websites from individual QICs

DATABASE SEARCHED & TIME PERIOD COVERED:

PubMed, prior to October 2015
Google, prior to October 2015

LANGUAGE:

English

SEARCH TERMS:

Official QIC name or name listed in NSQIP records	Alternative names used in search	Key personnel
Canadian National Surgical Quality Improvement Collaborative	Surgical Quality Action Network	Marsden
Connecticut Surgical Quality Coalition	CSQC, CtSQC, "Connecticut collaborative surgery NSQIP"	Lincer RM, Corvo PR
Florida Surgical Care Initiative	FSCI, Florida surgical care collaborative, FSCC, "Florida collaborative surgery NSQIP"	
Georgia Surgical Quality Collaborative	"Georgia collaborative surgery NSQIP"	Sweeney JF, Wilson B
Illinois Surgical Quality Improvement Collaborative	ISQIC, "Illinois collaborative surgery"	Bilimoria KY, O'Leary KJ, Yang AD, Johnson JK
Michigan Surgical Quality Collaborative	MSQC, "Michigan collaborative surgery"	Campbell DA, Englesbe MJ, Dimick JB
Pennsylvania NSQIP Consortium	"Pennsylvania collaborative surgery NSQIP"	
Surgical Care and Outcomes Assessment Program	SCOAP, SCOAP-CERTAIN	Thirlby RC, Kolios-Morris VA, Flum DR
Tennessee Surgical Quality Collaborative	TSQC, "Tennessee collaborative surgery NSQIP"	Cofer JB, Guillamondegui OD
Upstate New York Surgical Quality Initiative	UNYSQI, "New York collaborative surgery NSQIP"	Rickles AS, Cooney R, Vielhauer M

Appendix 2.4: Supplemental list of NSQIP QICs and reasons for exclusion

Name listed in NSQIP records	Year formed	Status in review	Reason for exclusion
Michigan Surgical Quality Collaborative	2005	Included, at least one publication found	
Partners HealthCare	2005	Excluded	Single system
Tennessee Surgical Quality Collaborative	2009	Included, at least one publication found	
Canadian National Surgical Quality Improvement Collaborative	2011	Included, at least one publication found	
Connecticut Surgical Quality Coalition	2011	Included, at least one publication found	
Department of Defense	2011	Excluded	Not regional
Florida Surgical Care Initiative	2011	Included, at least one publication found	
Kaiser Permanente Northern California Regional NSQIP Collaborative	2011	Excluded	Single system
Mayo Clinic Surgical Quality Consortium	2011	Excluded	Single system
Northern California Surgical Quality Collaborative	2011	Included, no publications found	
Nebraska Collaborative	2011	Included, no publications found	
Oregon NSQIP Consortium	2011	Included, no publications found	
Pennsylvania NSQIP Consortium	2011	Included, at least one publication found	
Surgical Quality Action Network	2011	Excluded	Part of another QIC
MaineHealth Collaborative	2012	Excluded	Single system

Upstate New York Surgical Quality Initiative	2012	Included, at least one publication found	
Virginia Surgical Quality Collaborative	2012	Included, no publications found	
Carolinas HealthCare System Surgery Quality & Safety Operations Council Collaborative	2013	Excluded	Single system
Fraser Health Systems	2013	Excluded	Single system
Georgia Surgical Quality Collaborative	2013	Included, at least one publication found	
HIC Surgical Safety Collaborative	2013	Excluded	Single system
Ontario Collaborative	2013	Included, no publications found	
University of California CHQI Collaborative	2013	Excluded	Single site for general surgery, only multisite for cardiac
Weill Cornell NYHQ NSQIP Collaborative	2013	Excluded	Single system
Covenant Health Collaborative	2014	Excluded	Single system
Illinois Surgical Quality Improvement Collaborative	2014	Included, no publications found	
University of Colorado Health Surgical Quality Collaborative	2014	Excluded	Single system
ACS NSQIP HPB Collaborative	2014	Excluded	Not regional
Memorial Hermann Healthcare System NSQIP Collaborative	2015	Excluded	Single system

Appendix 4.1: Crosswalk of MeSH terms from core set of QIC articles

Article	Term #1	Term #2	Term #3	Additional terms	Excluded terms[a]
Plsek 1999	Cooperative Behavior	Total Quality Management/methods*	Organizational Innovation*	Process Assessment (Health Care); Models, Organizational; Clinical Medicine/standards*; Clinical Medicine/organization & administration*; Planning Techniques	Data Collection/methods
Øvretveit et al. 2002	Cooperative Behavior*	Quality Assurance, Health Care/organization & administration* Quality Assurance, Health Care/methods	Organizational Objectives	Health Care Coalitions/organization & administration*; Diffusion of Innovation; Management Quality Circles	Humans, Sweden, Guidelines as Topic, Health Services Research
Wilson et al. 2003	Cooperative Behavior*	Total Quality Management/organization & administration*	Organizational Innovation	Health Care Coalitions/organization & administration*; Models, Organizational*; Interinstitutional Relations; Leadership; Program Development	Developed Countries, Humans, Interviews as Topic
Mills and Weeks 2004	Cooperative Behavior	Total Quality Management/methods*	Organizational Objectives		Analysis of Variance, Humans, United States, Surveys and Questionnaires, Hospitals, Veterans/standards*
Ayers et al. 2005	Cooperative Behavior*	Quality Assurance, Health Care*		Learning*	Interviews as Topic, Europe, Humans, United States
Dückers et al. 2008, Dückers et al. 2009	Cooperative Behavior	Quality Assurance, Health Care/organization & administration*	Organizational Culture*	Patient Care Team/organization & administration*; Management Audit;	Surveys and Questionnaires*, Humans, Netherlands, Psychometrics/instrumentation* , Health Services Research, Reproducibility of Results

Nembhard 2009	Cooperative Behavior*	Quality Assurance, Health Care/organization & administration*	Learning*; Institutional Management Teams*; Consumer Behavior*	Interviews as Topic, Canada, Humans, United States, Research Design, Cross-Sectional Studies, Health Services Research
Campbell 2009, Campbell et al. 2010	Cooperative Behavior*	Quality Assurance, Health Care* Quality Assurance, Health Care/trends*	Quality Indicators, Health Care*; Societies, Medical; Communication; Cost Control; Health Care Reform	Follow-Up Studies; Humans; Michigan; United States; Retrospective Studies; Surgical Procedures, Operative/standards*
Schouten et al. 2010				[b]
Hulscher et al. 2013	Cooperative Behavior*	Quality Improvement*	Outcome and Process Assessment (Health Care)*; Health Knowledge, Attitudes, Practice*; Patient Care Team/standards*	Humans
Nadeem et al. 2013	Cooperative Behavior*	Quality Improvement/standards*	Outcome Assessment (Health Care); Quality of Health Care/standards*	Humans
Ovretveit 2013				[c]
Carter 2014	Cooperative Behavior*	Quality Improvement/standards*		Humans, England, Qualitative Research, Stroke/therapy*

[a]Terms specific to site, methods, or clinical topic
[b]No MeSH terms listed
[c]Not indexed on PubMed

Appendix 4.2: Search strategy to confirm core set of QIC articles

DATABASE SEARCHED & TIME PERIOD COVERED:

PubMed, prior to April 6, 2016

LANGUAGE:

English

SEARCH STRATEGY:

(quality[tiab] AND improvement[tiab] AND collaborative[tiab]) OR ((series OR project) AND breakthrough) OR "Health Care Coalitions"[Mesh]

AND

("Cooperative Behavior"[Mesh]

AND

("Quality Assurance, Health Care"[Mesh] OR "Total Quality Management"[Mesh] OR "Quality Improvement"[Mesh])

AND

("Organizational Innovation"[Mesh] OR "Organizational Objectives"[Mesh] OR "Organizational Culture"[Mesh]))

OR

"Models, Organizational"[Mesh]

TOTAL: 301

Appendix 5.1: Stepwise Construction of Lasso Model

Step	Variable Entered[a]	Variable Removed[a]	AICC	Result
1	At least one reactive pupil (Yes/No)	-	325.47	Add
2	Cerebral edema on head CT	-	324.22	Add
3	Injury Severity Score	-	305.25	Add
4	Glasgow Coma Scale score	-	284.75	Add
5	Loss of basal cisterns on head CT	-	255.01	Add
6	Age = 76-85[b]	-	247.89	Add
7	Age = 86+[b]	-	248.54	Stop

AICC, Akaike Information Criteria (Corrected); CT, computed tomography
[a]Lasso model works by attempting to add or remove a variable in each step. For our data, the model added variables until it reached the optimal AICC value and did not remove any previously added variables.
[b]SAS PROC GLMSELECT requires dummy coding. Therefore, the variable for each age group was eligible to be added or removed from the model separately.

Appendix 5.2: Comparison of Regression Coefficients Across Hierarchical Logistic, Ordinary Logistic, and Lasso Models

Variable	Hierarchical logistic model with "shrinkage"	Ordinary "unshrunk" logistic model	Lasso penalized model[a]
		Beta coefficient (95% CI)	
At least one reactive pupil (Yes/No)	-1.47***	-1.43***	0.182
	(-1.94, -1.01)	(-1.88, -0.98)	
Age, years (ref = 18-25)			
26-35	0.68	0.69	
	(-0.05, 1.41)	(-0.04, 1.41)	
36-55	1.07**	1.08**	
	(0.43, 1.71)	(0.44, 1.71)	
56-65	2.15***	2.17***	
	(1.37, 2.94)	(1.39, 2.95)	
66-75	2.75***	2.72***	
	(1.84, 3.65)	(1.83, 3.62)	
76-85	3.22***	3.20***	0.031
	(2.31, 4.13)	(2.3, 4.1)	

86+	4.13***	4.08***	
	(2.98, 5.27)	(2.95, 5.22)	
Injury Severity Score	0.07***	0.07***	0.003
	(0.05, 0.09)	(0.05, 0.09)	
Cerebral edema on head CT	1.73***	1.73***	0.125
	(1.25, 2.21)	(1.26, 2.21)	
Glasgow Coma Scale score	-0.25***	-0.25***	-0.015
	(-0.38, -0.12)	(-0.38, -0.12)	
Loss of basal cisterns on head CT	1.30***	1.37***	0.062
	(0.73, 1.87)	(0.84, 1.91)	
Intercept	-2.80***	-2.69***	-0.283
	(-3.88, -1.71)	(-3.73, -1.65)	
Random intercept (Center)	0.06		
	(0.003, 1.35)		

* p<0.05, ** p<0.01, *** p<0.001
[a]Results from linear LASSO penalized model using SAS PROC GLMSELECT (Died=1, Did not die=0). Procedure does not allow for the construction of confidence intervals or significance testing

Appendix 5.3: Distribution of Variables by Subsample

Variable	Frequency in development sample	Frequency in validation sample	P-value for comparison between samples[a]
Died	41.97	38.79	0.379
Age (years)			0.343[b]
18-25	21.13	21.11	
26-35	18.87	15.04	
36-55	29.86	29.82	
56-65	12.39	13.46	
66-75	7.32	7.92	
76-85	5.07	8.97	
86+	5.35	3.69	
Female	24.23	24.01	0.946
Race			0.226[b,c]
Hispanic	43.10	37.73	
White	30.42	38.52	
Black	11.55	12.40	

Asian	9.58	7.65	
Other	4.51	2.90	
Unknown	0.85	0.79	
Mechanism of Injury			0.854[b,c]
Fall	30.70	33.25	
Auto vs. Pedestrian	28.45	29.02	
Assault	13.80	12.93	
Motor Vehicle Collision	11.55	12.40	
Motorcycle Collision	8.73	7.39	
Other	6.76	5.01	
Injury Severity Score, mean (SD)	29.28 (12.97)	28.88 (12.02)	0.666[c]
Glasgow Coma Scale, mean (SD)	4.48 (1.86)	4.41 (1.86)	0.612
Heart Rate (beats per minute)[d]			0.146[b]
0-89	40.28	33.51	
90-119	36.90	39.58	
120+	22.82	26.91	
Systolic Blood Pressure (mm Hg)[d]			0.607[b]
0-99	8.73	11.35	
100-149	49.86	50.13	
150-199	33.52	30.34	
200+	7.89	8.18	
Hypertension	17.75	20.84	0.289
Alcoholism	13.24	12.40	0.734
Diabetes	6.20	11.87	0.009
Current Smoker	4.51	6.33	0.278
Obesity	2.25	1.85	0.698
Respiratory Disease	1.69	2.11	0.678
Heart Disease	1.13	1.85	0.427
Cerebral edema on head CT	25.07	28.50	0.296
At least one reactive pupil (Yes/No)	67.04	64.38	0.448
Loss of basal cisterns on head CT	18.59	18.47	0.966

CT, computed tomography; SD, standard deviation
[a]Based on Wald test from univariate hierarchical logistic regression model
[b]Based on overall F-test instead of Wald test as variable had multiple levels
[c]Hierarchical model did not converge; result from 200th iteration
[d]Includes imputed data

Appendix 5.4: Frequency of Comorbidities by Subsample

Variable	Overall sample	Development sample (n=355)			Validation sample (n=379)		
		Subsample	Did not die	Died	Subsample	Did not die	Died
Hypertension	142	63	32	31	79	40	39
Alcoholism	94	47	35	12	47	35	12
Diabetes	67	22	6	16	45	24	21
Current Smoker	40	16	10	6	24	20	4
Obesity	15	8	2	6	7	6	1
Respiratory Disease	14	6	4	2	8	4	4
Heart Disease	11	4	3	1	7	2	5
Bleeding Disorder	9	4	0	4	5	3	2
Functionally Dependent	8	2	0	2	6	4	2
Pre-hospital Cardiac Arrest	8	2	0	2	6	0	6
Stroke	5	1	1	0	4	1	3
Cancer	4	0	0	0	4	1	3
Liver Disease	3	3	1	2	0	0	0

All numbers represent frequencies
Light shading indicates variables that perfectly predicted the outcome in at least one subsample
Dark shading indicates variables with frequency of zero in at least one subsample

Appendix 5.5: Correlation Matrix for All Potential Covariates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	1												
(2)	-0.20	1											
(3)	-0.08	0.10	1										
(4)	-0.31	0.14	0.07	1									
(5)	-0.11	-0.04	-0.01	0.12	1								
(6)	0.01	0.08	0.01	0.06	-0.30	1							
(7)	-0.10	-0.01	0.05	0.10	0.10	0.01	1						
(8)	0.20	0.06	-0.03	-0.10	-0.18	0.09	-0.07	1					
(9)	0.24	-0.03	0.05	-0.14	-0.05	0.05	-0.09	0.11	1				
(10)	0.09	0.04	-0.06	0.00	-0.03	0.04	-0.04	0.03	0.12	1			
(11)	0.06	0.10	-0.03	-0.05	-0.12	0.07	0.02	0.03	0.02	0.02	1		
(12)	0.43	-0.14	0.02	-0.27	-0.11	-0.03	-0.08	0.17	0.42	0.05	0.02	1	
(13)	0.04	0.01	0.03	-0.04	-0.01	-0.04	-0.02	0.04	0.05	-0.02	0.09	0.12	1
(14)	0.04	-0.04	-0.01	0.05	0.00	-0.02	0.04	0.01	0.06	-0.02	0.07	0.03	0.05
(15)	-0.01	0.05	-0.05	0.02	-0.01	-0.02	0.03	0.00	0.01	-0.03	0.12	0.00	0.05
(16)	-0.03	0.12	0.06	-0.01	-0.23	0.32	0.04	0.03	0.02	-0.05	0.07	0.06	-0.02
(17)	0.07	-0.11	-0.01	-0.01	0.17	-0.12	0.04	-0.08	0.01	0.03	-0.06	0.03	0.07
(18)	0.12	0.02	0.01	-0.07	0.06	-0.06	-0.15	0.05	0.09	0.09	0.01	0.03	-0.02
(19)	0.03	-0.03	0.05	0.00	0.10	-0.05	-0.04	-0.06	-0.04	-0.01	-0.01	-0.07	0.00
(20)	-0.06	0.09	0.01	0.09	-0.01	0.13	-0.03	-0.04	-0.09	0.00	0.01	-0.01	0.04
(21)	-0.10	0.00	0.01	0.04	0.23	-0.18	0.01	-0.06	-0.05	0.00	-0.10	-0.03	0.00
(22)	-0.04	0.00	0.06	0.03	0.00	-0.01	0.04	-0.07	0.01	-0.04	-0.03	-0.04	0.06
(23)	0.07	-0.05	-0.04	-0.04	0.03	-0.06	-0.05	0.06	0.04	0.02	-0.01	0.02	0.00
(24)	0.07	-0.10	0.05	0.01	0.14	-0.16	-0.08	0.03	-0.01	0.03	-0.13	-0.03	-0.02
(25)	-0.12	0.08	0.04	0.01	-0.03	-0.04	-0.03	0.02	-0.09	-0.04	-0.01	-0.06	-0.02
(26)	-0.09	-0.06	-0.01	0.03	0.16	-0.15	0.05	-0.14	0.01	0.01	-0.08	-0.05	-0.04

	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
(14)	1												
(15)	0.05	1											
(16)	-0.05	0.02	1										
(17)	-0.02	-0.09	-0.19	1									
(18)	0.00	-0.08	-0.09	0.00	1								
(19)	-0.03	0.00	0.00	0.01	0.00	1							
(20)	-0.02	0.01	0.17	-0.07	-0.11	0.13	1						
(21)	-0.04	0.00	-0.20	0.12	0.09	0.13	0.08	1					
(22)	-0.01	-0.02	-0.02	0.03	-0.13	0.08	0.10	0.06	1				
(23)	0.00	-0.03	-0.06	0.13	0.18	-0.08	0.05	0.14	0.02	1			
(24)	0.01	-0.05	-0.25	0.08	0.21	0.12	0.03	0.21	0.01	0.17	1		
(25)	0.02	0.04	0.05	-0.08	0.04	-0.03	-0.01	-0.03	-0.04	0.05	-0.01	1	
(26)	0.00	0.04	-0.19	0.15	0.04	0.09	-0.06	0.35	0.04	0.11	0.21	-0.01	1

(1) Age, (2) Female, (3) Race, (4) Mechanism of Injury, (5) Injury Severity Score, (6) Glasgow Coma Scale, (7) Heart Rate, (8) Systolic Blood Pressure, (9) Diabetes, (10) Heart Disease, (11) Alcoholism, (12) Hypertension, (13) Obesity, (14) Respiratory Disease, (15) Current Smoker, (16) Reactive Pupil, (17) Elevated INR, (18) Subdural Hematoma, (19) Subarachnoid Hemorrhage, (20) Intraparenchymal Contusion, (21) Cerebral Edema, (22) Intracranial Hematoma, (23) Mass Effect, (24) Loss of Basal Cisterns, (25) Epidural Hematoma, (26) Loss of Grey/White Differential

Appendix 5.6: Exploring Sources of Overfitting in the *And* Model

Variable	Beta coefficient in development sample (95% CI)	Beta coefficient in validation sample (95% CI)
Age, years (ref = 18-25)		
26-35	0.87 (-0.33, 2.07)	0.78 (-0.5, 2.06)
36-55	0.45 (-0.67, 1.57)	1.47** (0.38, 2.56)
56-65	1.91** (0.54, 3.29)	2.06* (0.77, 3.35)
66-75	3.17**	2.21**

	(1.25, 5.1)	(0.64, 3.77)
76-85	4.36***	2.91***
	(2.25, 6.47)	(1.33, 4.49)
86+	4.03***	3.57**
	(1.96, 6.1)	(1.42, 5.72)
Female	0.60	-0.97*
	(-0.35, 1.55)	(-1.74, -0.19)
Race (ref = White)		
Black	-0.90	0.51
	(-2.19, 0.38)	(-0.65, 1.67)
Hispanic	-0.97*	1.19**
	(-1.94, 0)	(0.32, 2.06)
Other	-1.24	0.96
	(-3.02, 0.55)	(-1.35, 3.26)
Unknown	-0.40	1.69
	(-3.94, 3.14)	(-2.25, 5.63)
Asian/Pacific Islander	0.08	1.24*
	(-1.35, 1.5)	(0.04, 2.44)
Mechanism of Injury (ref = Fall)		
Auto vs. Pedestrian	0.23	-1.10*
	(-0.81, 1.28)	(-2.18, -0.02)
Motorcycle Collision	-0.95	-2.10*
	(-2.59, 0.69)	(-3.88, -0.33)
Motor Vehicle Collision	-1.83*	-1.56*
	(-3.31, -0.35)	(-2.92, -0.2)
Assault	-0.20	-1.69*
	(-1.48, 1.08)	(-3.00, -0.38)
Other	0.58	-0.18
	(-0.97, 2.13)	(-1.72, 1.35)
Injury Severity Score, mean (SD)	0.08***	0.10***
	(0.05, 0.12)	(0.06, 0.15)
Glasgow Coma Scale, mean (SD)	-0.35**	-0.27*
	(-0.59, -0.12)	(-0.48, -0.06)
Heart Rate, beats per minute (ref = 0-89)		
90-119	0.67	0.07
	(-0.17, 1.52)	(-0.74, 0.87)
120+	0.04	0.03
	(-0.92, 1.01)	(-0.88, 0.94)
Systolic Blood Pressure, mm Hg (ref = 0-99)		
100-149	-0.79	-0.39

	(-2.08, 0.5)	(-1.49, 0.72)
150-199	-0.19	0.43
	(-1.55, 1.17)	(-0.76, 1.63)
200+	0.50	1.13
	(-1.27, 2.27)	(-0.47, 2.72)
Diabetes	2.70**	0.63
	(0.95, 4.44)	(-0.5, 1.75)
Heart Disease	-0.69	1.10
	(-3.36, 1.99)	(-1.36, 3.56)
Alcoholism	-1.00	0.13
	(-2.22, 0.21)	(-0.97, 1.24)
Hypertension	-1.21*	-0.46
	(-2.4, -0.02)	(-1.46, 0.55)
Obesity	3.08*	-2.70
	(0.52, 5.65)	(-5.86, 0.46)
Respiratory Disease	-1.58	1.04
	(-4.5, 1.35)	(-1.43, 3.52)
Current Smoker	0.25	-1.73*
	(-1.57, 2.08)	(-3.33, -0.13)
Cerebral edema on head CT	2.17**	1.89***
	(1.31, 3.03)	(1.06, 2.72)
At least one reactive pupil (Yes/No)	-1.64***	-1.78***
	(-2.48, -0.81)	(-2.61, -0.96)
Loss of basal cisterns on head CT	1.31**	1.43**
	(0.36, 2.26)	(0.54, 2.31)
Intercept	-2.34	-1.87
	(-5.25, 0.57)	(-4.17, 0.42)

* p<0.05, ** p<0.01, *** p<0.001
Shading represents variables for which significant testing was not the same across the two samples

Variations of the *And* model	AUROC (95% CI[a]) Development sample (2010 data)	AUROC (95% CI[a]) Validation sample (2011 data)
Full model	0.942 (0.917, 0.962)	0.847 (0.801, 0.881)
Without diabetes	0.938 (0.911, 0.960)	0.866 (0.828, 0.902)
Without diabetes and current smoker	0.938 (0.910, 0.957)	0.867 (0.827, 0.903)
Without diabetes, current smoker, and gender	0.937 (0.912, 0.958)	0.870 (0.829, 0.900)

AUROC, area under the receiver operator curve; CI, confidence interval
[a]95% confidence intervals bootstrapped and bias-corrected with 1,000 repetitions

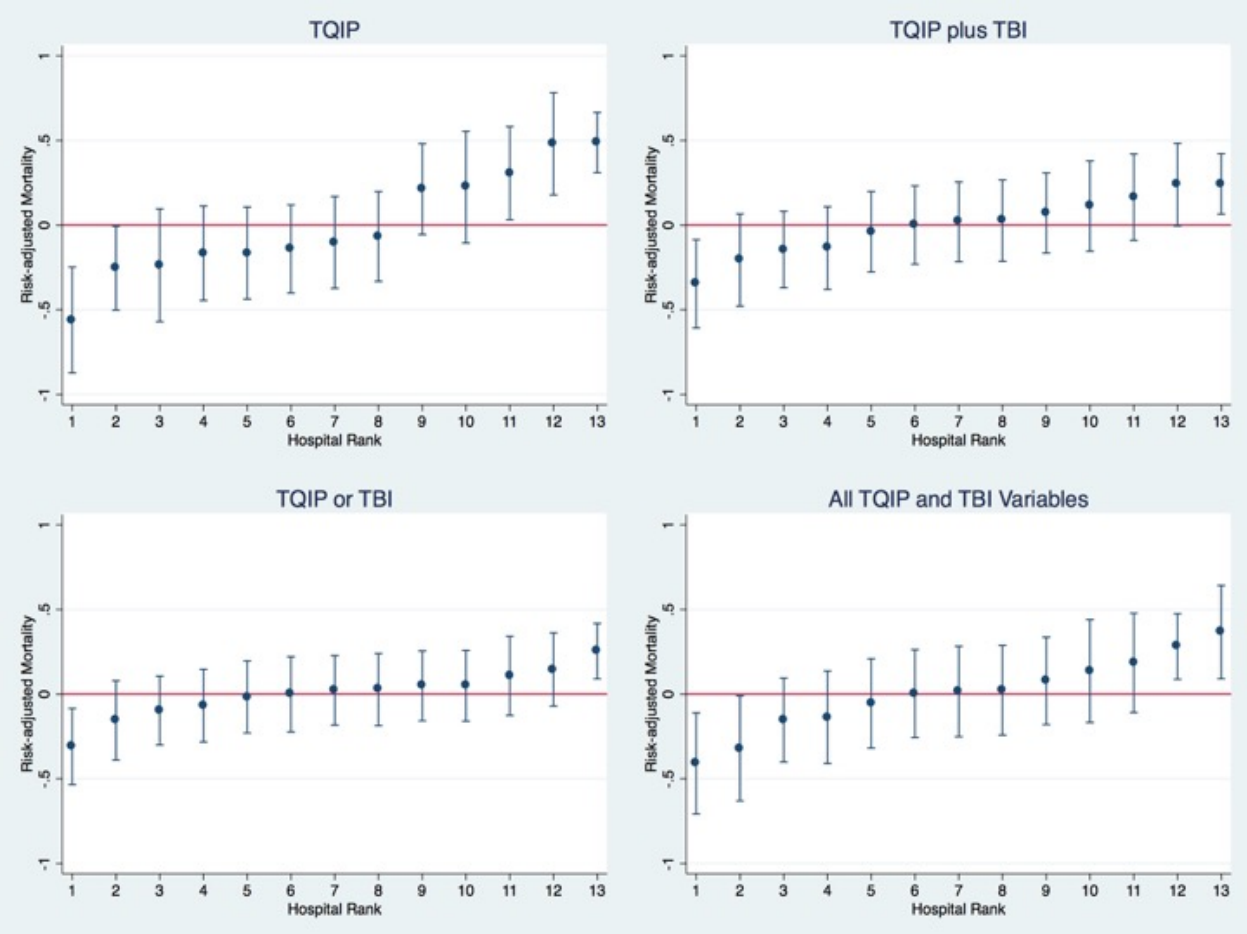
Appendix 5.7: Full Regression Coefficients from All Four Model Specifications

Variable	Beta coefficient in *Standard TQIP* model	Beta coefficient in *And* model	Beta coefficient in *Or* model	Beta coefficient in *All variables* model
Age (years)				
26-35	0.72*	0.64	0.68	0.78
36-55	0.49	0.90*	1.07**	0.99**
56-65	1.43***	1.82***	2.15***	1.92***
66-75	1.90***	2.18***	2.75***	2.28***
76-85	2.18***	2.80***	3.22***	2.78***
86+	2.90***	3.48***	4.13***	3.33***
Female	-0.56*	-0.4		-0.31
Race (ref = White)				
Black	0.01	0.07		0.06
Hispanic	0.42	0.42		0.47
Other	-0.18	-0.09		-0.34
Unknown	0.55	0.52		0.34
Asian	0.43	0.84*		1.10*
Mechanism of Injury (ref = Fall)				
Auto vs. Pedestrian	-0.57*	-0.55		-0.4
Motorcycle Collision	-1.64***	-1.57**		-1.34*
Motor Vehicle Collision	-1.73***	-1.79***		-1.80***
Assault	-0.81*	-1.06*		-1.00*
Other	-0.15	-0.37		-0.49
Injury Severity Score, mean (SD)	0.10***	0.09***	0.07***	0.09***
Glasgow Coma Scale, mean (SD)	-0.42***	-0.27***	-0.25***	-0.29***
Heart Rate, beats per minute (ref = 0-89)				
90-119	-0.05	0.16		0.22
120+	-0.26	-0.03		-0.1
Systolic Blood Pressure, mm Hg (ref = 0-99)				
100-149	-0.68	-0.64		-0.51
150-199	-0.15	-0.04		0.14
200+	0.97*	0.72		0.96
Diabetes	0.54	0.79		0.47
Heart Disease	1.06	0.27		0.13
Alcoholism	-0.82*	-0.36		-0.43
Hypertension	-0.73*	-0.67		-0.57
Obesity	0.11	0.28		0.35

Respiratory Disease	0.29	-0.11		0.02
Current Smoker	-1.08*	-1.1		-1.04
Cerebral edema on head CT		1.88***	1.73***	1.68***
At least one reactive pupil (Yes/No)		-1.48***	-1.47***	-1.34***
Loss of basal cisterns on head CT		1.22***	1.30***	1.03**
Elevated INR (>1.4)				0.79
Subdural Hematoma				0.4
Subarachnoid Hemorrhage				-0.3
Intraparenchymal Contusion				-0.06
Intracranial Hematoma				0.07
Mass Effect				0.47
Epidural Hematoma				-1.21**
Loss of Grey/White Differential				1.42**
Intercept	-0.55	-1.55	-2.80***	-2.18*
Random intercept (Center)	0.17	0.09	0.06	0.12

* p<0.05, ** p<0.01, *** p<0.001

Appendix 5.8: Caterpillar Plots for All Four Model Specifications



Appendix 6.1: List of International Classification of Diseases, Ninth Edition Procedural Codes for Craniotomy

ICD-9 Code	Description
01.2	Craniotomy and craniectomy
01.24	Other craniotomy
01.25	Other craniectomy
01.31	Incision of cerebral meninges
01.53	Lobectomy of brain

Appendix 6.2: Comparison of Independently-Abstracted Variable for ICP Monitor Placement with Recorded *International Classification of Diseases, Ninth Edition* Procedural Code

ICD-9 Code	Description
1.10	Intracranial pressure monitoring
1.16	Intracranial oxygen monitoring
1.17	Brain temperature monitoring
1.18	Other diagnostic procedures on brain and cerebral meninges
2.20	Ventriculostomy

ICP Monitor Placement		ICD-9 Coding		
		No	Yes	Total
Registry variable	No	427	17	444
	Yes	43	335	378
	Total	470	352	822

Cohen's $\kappa = 0.8523$

p-value <0.001

Percent agreement = 92.7%

Cohen's level of agreement = very good

Appendix 6.3: Comparison of Regression Coefficients Between Ordinary Logistic and Hierarchical Logistic Regression Models

Variable	Beta coefficient from ordinary "unshrunk" logistic model (95% CI)	Beta coefficient from hierachical logistic model with "shrinkage" (95% CI)
Age, years (ref = 18-25)		
26-35	0.78	0.78
	(-0.07, 1.63)	(-0.09, 1.65)
36-55	0.96*	0.95*
	(0.22, 1.69)	(0.20, 1.70)
56-65	1.92***	1.91***
	(1.02, 2.83)	(0.98, 2.84)
66-75	2.18***	2.2***
	(1.09, 3.26)	(1.08, 3.31)
76-85	2.75***	2.82***
	(1.09, 3.26)	(1.67, 3.96)
86+	3.19***	3.19***
	(1.79, 4.59)	(1.76, 4.62)
Female	-0.30	-0.28
	(-0.86, 0.27)	(-0.85, 0.29)
Race (ref = White)		
Black	0.03	0.04
	(-0.80, 0.86)	(-0.82, 0.90)
Hispanic	0.52	0.41
	(-0.08, 1.11)	(-0.22, 1.04)
Other	-0.12	-0.42
	(-1.45, 1.21)	(-1.89, 1.04)
Unknown	0.20	0.31
	(-2.10, 2.51)	(-2.04, 2.66)
Asian/Pacific Islander	1.06**	1.08*
	(0.21, 1.91)	(0.19, 1.97)
Mechanism of Injury (ref = Fall)		
Auto vs. Pedestrian	-0.34	-0.44
	(-0.99, 0.31)	(-1.13, 0.24)
Motorcycle Collision	-1.35*	-1.39*

	(-2.42, -0.29)	(-2.49, -0.29)
Motor Vehicle Collision	-1.73***	-1.88***
	(-2.67, -0.79)	(-2.87, -0.88)
Assault	-0.94*	-1.04*
	(-1.79, -0.1)	(-1.92, -0.15)
Other	-0.46	-0.55
	(-1.49, 0.57)	(-1.60, 0.51)
Injury Severity Score, mean (SD)	0.08***	0.09***
	(0.06, 0.11)	(0.07, 0.12)
Glasgow Coma Scale, mean (SD)	-0.29***	-0.30***
	(-0.44, -0.14)	(-0.45, -0.14)
Heart Rate,beats per minute (ref = 0-89)		
90-119	0.23	0.18
	(-0.3, 0.77)	(-0.37, 0.74)
120+	-0.01	-0.09
	(-0.64, 0.62)	(-0.75, 0.57)
Systolic Blood Pressure, mm Hg (ref = 0-99)		
100-149	-0.50	-0.49
	(-1.33, 0.31)	(-1.33, 0.35)
150-199	0.14	0.20
	(-0.72, 1.00)	(-0.69, 1.09)
200+	0.96	1.06
	(-0.20, 2.11)	(-0.13, 2.25)
Hypertension	-0.58	-0.62
	(-1.28, 0.11)	(-1.33, 0.09)
Alcoholism	-0.36	-0.31
	(-1.13, 0.41)	(-1.11, 0.48)
Diabetes	0.54	0.47
	(-0.31, 1.40)	(-0.42, 1.35)
Current smoker	-1.11	-1.00
	(-2.24, 0.02)	(-2.17, 0.17)
Obesity	0.38	0.29
	(-1.27, 2.02)	(-1.39, 1.98)
Respiratory disease	0.29	0.30
	(-1.53, 2.10)	(-1.55, 2.15)
Heart disease	0.26	0.16
	(-1.69, 2.20)	(-1.80, 2.11)
Bleeding disorder	0.80	1.12

	(-1.46, 3.06)	(-1.24, 3.48)
Functionally dependent	1.27	1.51
	(-0.69, 3.22)	(-0.45, 3.48)
Stroke	-0.03	-0.26
	(-2.67, 2.61)	(-2.97, 2.45)
Cancer	0.75	0.96
	(-1.90, 3.40)	(-1.78, 3.70)
Liver disease	-1.88	-2.17
	(-5.02, 1.25)	(-5.38, 1.05)
At least one reactive pupil	-1.30***	-1.39***
	(-1.83, -0.78)	(-1.94, -0.84)
Elevated INR	0.92*	0.93*
	(0.10, 1.74)	(0.09, 1.77)
Subdural hematoma	0.48	0.47
	(-0.06, 1.01)	(-0.08, 1.02)
Subarachnoid hemorrhage	-0.25	-0.30
	(-0.76, 0.25)	(-0.82, 0.22)
Intraparenchymal contusion	-0.01	-0.07
	(-0.53, 0.51)	(-0.61, 0.48)
Cerebral edema	1.67***	1.67***
	(1.12, 2.23)	(1.09, 2.24)
Intracranial hematoma	0.08	0.05
	(-0.49, 0.66)	(-0.53, 0.64)
Mass effect	0.49	0.46
	(-0.11, 1.09)	(-0.16, 1.07)
Loss of basal cisterns	1.09**	1.01**
	(0.45, 1.72)	(0.35, 1.68)
Epidural hematoma	-1.17**	-1.22**
	(-2.01, -0.32)	(-2.08, -0.36)
Loss of grey/white differential	1.38*	1.45**
	(0.32, 2.45)	(0.36, 2.54)
Intercept	-2.25*	-2.43**
	(-3.97, -0.53)	(-4.23, -0.62)
Random intercept (Center)		0.25
		(0.03, 2.11)

* p<0.05, ** p<0.01, *** p<0.001
Shading represents variables for which significant testing was not the same across the two models

Appendix 6.4: Comparison of Hospitals' Observed-to-Expected Ratios Between Ordinary Logistic and Hierarchical Logistic Regression Models

Hospital[a]	Observed-to-Expected Mortality Ratio[b]	
	Ordinary "unshrunk" logistic model (95% CI[c])	Hierarchical logistic model with "shrinkage"[d] (95% CI[c])
A	0.68* (0.47, 0.95)	0.56* (0.26, 0.89)
B	0.74* (0.53, 0.99)	0.66* (0.36, 0.97)
C	0.97 (0.93, 1.05)	0.92 (0.79, 1.11)
D	0.97 (0.90, 1.03)	0.94 (0.77, 1.08)
E	0.98 (0.90, 1.13)	0.95 (0.74, 1.20)
F	0.99 (0.89, 1.09)	0.98 (0.78, 1.21)
G	1.00 (0.88, 1.10)	0.98 (0.71, 1.27)
H	1.01 (0.95, 1.06)	1.00 (0.85, 1.16)
I	1.02 (1.00, 1.05)	1.08 (1.00, 1.19)
J	1.03 (0.92, 1.12)	1.05 (0.86, 1.31)
K	1.08 (0.97, 1.16)	1.13 (0.92, 1.33)
L	1.11 (0.93, 1.26)	1.17 (0.86, 1.51)
M	1.13* (1.01, 1.19)	1.25* (1.01, 1.58)

[a]Codes are intentionally arbitrary to protect hospitals' identities.
[b]Observed-to-expected ratios were calculated by dividing the number of observed deaths by the number of predicted deaths under the regression model.
[c]All 95% CI values represent percentile confidence intervals generated via bootstrapping with 1,000 repetitions.
[d]Hierarchical models control for clustering with hospital-level random intercepts.

References

1. National Library of Medicine, National Institutes of Health. “Quality of Health Care”[Mesh]. pubmed.org. http://www.ncbi.nlm.nih.gov/pubmed?term=%22Quality+of+Health+Care%22%5BMesh%5D. Accessed June 10, 2015.

2. Casalino LP, Gans D, Weber R, et al. US Physician Practices Spend More Than $15.4 Billion Annually To Report Quality Measures. *Health Aff (Millwood)*. 2016;35(3):401-406. doi:10.1377/hlthaff.2015.1258.

3. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *N Engl J Med*. 1996;335(13):966-970. doi:10.1056/NEJM199609263351311.

4. Blumenthal D. Quality of health care. Part 4: The origins of the quality-of-care debate. *N Engl J Med*. 1996;335(15):1146-1149. doi:10.1056/NEJM199610103351511.

5. Porter ME, Lee TH. The strategy that will fix health care. *Harvard Business Review*. 2013.

6. Pryor D, Hendrich A, Henkel RJ, Beckmann JK, Tersigni AR. The quality 'journey‘ at Ascension Health: how we’ve prevented at least 1,500 avoidable deaths a year--and aim to do even better. *Health Affairs*. 2011;30(4):604-611. doi:10.1377/hlthaff.2010.1276.

7. Pronovost PJ, Marsteller JA, Goeschel CA. Preventing bloodstream infections: a measurable national success story in quality improvement. *Health Affairs*. 2011;30(4):628-634. doi:10.1377/hlthaff.2011.0047.

8. Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *N Engl J Med*. 2009;360(5):491-499. doi:10.1056/NEJMsa0810119.

9. Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med*. 2014;370(11):1029-1038. doi:10.1056/NEJMsa1308261.

10. Audet A-MJ, Doty MM, Shamasdin J, Schoenbaum SC. Measure, learn, and improve: physicians' involvement in quality improvement. *Health Aff (Millwood)*. 2005;24(3):843-853. doi:10.1377/hlthaff.24.3.843.

11. Plsek PE. Collaborating across organizational boundaries to improve the quality of care. *Am J Infect Control*. 1997;25(2):85-95.

12. Donabedian A. *Explorations in Quality Assessment and Monitoring: the Definition of Quality and Approaches to Its Assessment*. 1982.

13. Brook RH, Williams KN. Quality of health care for the disadvantaged. *J Community Health*. 1975;1(2):132-156.

14. Wennberg J, Gittelsohn. Small area variations in health care delivery. *Science*. 1973;182(4117):1102-1108.

15. McGlynn EA, Asch S, 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. doi:10.1056/NEJMsa022615.

16. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs*. 2015;34(3):423-430. doi:10.1377/hlthaff.2014.0201.

17. Langley GJ, Moen R, Nolan KM, Nolan TW, Norman C, Provost L. *The Improvement Guide: a Practical Approach to Enhancing Organizational Performance*. 2nd ed. San Francisco: Jossey-Bass; 2009.

18. Berwick DM, Godfrey AB, Roessner J. *Curing Health Care: New Strategies for Quality Improvement*. Paperback. San Francisco: Jossey-Bass; 2002.

19. Khuri SF, Daley J, Henderson W, et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *Journal of the American College of Surgeons*. 1995;180(5):519-531.

20. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg*. 1998;228(4):491-507.

21. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994;271(10):761-766.

22. Fink AS, Campbell DA, Mentzer RM, et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Ann Surg*. 2002;236(3):344–53–discussion353–4. doi:10.1097/01.SLA.0000027082.79556.55.

23. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg*. 2012;256(6):973-981. doi:10.1097/SLA.0b013e31826b4c4f.

24. Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg*. 2009;250(3):363-376. doi:10.1097/SLA.0b013e3181b4148f.

25. Cohen ME, Liu Y, Ko CY, Hall BL. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. *Ann Surg*. 2016;263(2):267-273. doi:10.1097/SLA.0000000000001192.

26. Osborne NH, Nicholas LH, Ryan AM, Thumma JR, Dimick JB. Association of Hospital Participation in a Quality Reporting Program With Surgical Outcomes and Expenditures for Medicare Beneficiaries. *JAMA*. 2015;313(5):496-504. doi:10.1001/jama.2015.25.

27. Etzioni DA, Wasif N, Dueck AC, et al. Association of Hospital Participation in a Surgical Outcomes Monitoring Program With Inpatient Complications and Mortality. *JAMA*. 2015;313(5):505-511. doi:10.1001/jama.2015.90.

28. Osborne NH, Etzioni DA. Does Participation in the ACS-NSQIP Improve Outcomes? *Ann Surg*. July 2015. doi:10.1097/SLA.0000000000001343.

29. Coronado VG, Xu L, Basavaraju SV, et al. Surveillance for traumatic brain injury-related deaths--United States, 1997-2007. *MMWR Surveill Summ*. 2011;60(5):1-32.

30. Dawes AJ, Sacks GD, Cryer HG, et al. Intracranial pressure monitoring and inpatient mortality in severe traumatic brain injury: A propensity score-matched analysis. *J Trauma Acute Care Surg*. 2015;78(3):492-502. doi:10.1097/TA.0000000000000559.

31. Mittman BS. Creating the Evidence Base for Quality Improvement Collaboratives. *Ann Intern Med*. 2004;140(11):897-901. doi:10.7326/0003-4819-140-11-200406010-00011.

32. Kilo CM, Kabcenell A, Berwick DM. Beyond survival: toward continuous improvement in medical care. *New Horiz*. 1998;6(1):3-11.

33. Chassin MR. Assessing strategies for quality improvement. *Health Aff (Millwood)*. 1997;16(3):151-161.

34. Blumenthal D, Kilo CM. A report card on continuous quality improvement. *Milbank Q*. 1998;76(4):625–48–511.

35. Shortell SM, Bennett CL, Byck GR. Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q*. 1998;76(4):593–624–510.

36. Samsa G, Matchar D. Can continuous quality improvement be assessed using randomized trials? [see comment]. *Health Serv Res*. 2000;35(3):687-700.

37. O'Neill SM, Hempel S, Lim YW, et al. Identifying continuous quality improvement publications: what makes an improvement intervention 'CQI'? *BMJ Quality & Safety*. 2011;20(12):1011-1019. doi:10.1136/bmjqs.2010.050880.

38. Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med*. 1989;320(1):53-56. doi:10.1056/NEJM198901053200110.

39. Plsek PE. Techniques for managing quality. *Hosp Health Serv Adm*. 1995;40(1):50-79.

40. Plsek PE. Quality Improvement Methods in Clinical Medicine. *Pediatrics*. 1999;103(Supplement E1):203-214. doi:10.1056/NEJM198901053200110.

41. Laffel G, Blumenthal D. The case for using industrial quality management science in health care organizations. *JAMA*. 1989;262(20):2869-2873.

42. Berwick DM. Quality comes home. *Ann Intern Med*. 1996;125(10):839-843.

43. Pascale RT, Sternin J, Sternin M. *The Power of Positive Deviance: How Unlikely Innovators Solve the World's Toughest Problems*. Boston: Harvard Business Review Press; 2010.

44. Berwick DM. The science of improvement. *JAMA: The Journal of the American Medical Association*. 2008;299(10):1182-1184. doi:10.1001/jama.299.10.1182.

45. Berwick DM. Eleven worthy aims for clinical leadership of health system reform. *JAMA*. 1994;272(10):797-802.

46. Haddock CC, Nosky C, Fargason CA, Kurz RS. The impact of CQI on human resources management. *Hosp Health Serv Adm*. 1995;40(1):138-153.

47. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Quarterly*. 2011;89(2):167-205. doi:10.1111/j.1468-0009.2011.00625.x.

48. The Joint Commission. About The Joint Commission. jointcommission.org. http://www.jointcommission.org/about_us/about_the_joint_commission_main.aspx. Published 2016. Accessed February 29, 2016.

49. Barsness ZI, Shortell SM, Gillies RR, et al. National Survey of Hospital Quality Improvement Activities. *Hospital Health Networks*. 1993;(5 December):52-55.

50. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ*. 1998;317(7156):465-468.

51. Lindenauer PK. Effects of Quality Improvement Collaboratives. *BMJ : British Medical Journal*. 2008;336(7659):1448-1449. doi:10.2307/20510129?ref=no-x-route:70089ef3cbcda172510ebbc5760a2d1c.

52. Shortell SM, O'Brien JL, Carman JM, et al. Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. *Health Serv Res*. 1995;30(2):377-401.

53. D'Andreamatteo A, Ianni L, Lega F, Sargiacomo M. Lean in healthcare: A comprehensive review. *Health policy*. February 2015. doi:10.1016/j.healthpol.2015.02.002.

54. Nicolay CR, Purkayastha S, Greenhalgh A, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical

healthcare. *Br J Surg*. 2011;99(3):324-335. doi:10.1002/bjs.7803.

55. Mason SE, Nicolay CR, Darzi A. The use of Lean and Six Sigma methodologies in surgery: a systematic review. *Surgeon*. 2015;13(2):91-100. doi:10.1016/j.surge.2014.08.002.

56. Shortell SM, Levin DZ, O'Brien JL, Hughes EFX. Assessing the evidence on CQI: is the glass half empty or half full? *Hosp Health Serv Adm*. 1995;40(1):4-24.

57. O'Brien JL, Shortell SM, Hughes EFX, et al. An integrative model for organization-wide quality improvement: lessons from the field. *Quality Management in Healthcare*. 1995;3(4):19-30.

58. Institute for Healthcare Improvement. *The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement*. Boston: Institute for Healthcare Improvement; 2003. doi:10.2337/diaspect.17.2.97.

59. Batalden PB. Personal communication. March 2016.

60. Kilo CM. A Framework for Collaborative Improvement: Lessons from the Institute for Healthcare Improvement's Breakthrough Series. *Quality Management in Healthcare*. 1998;6(4):1-13.

61. Kilo CM. Improving care through collaboration. *Pediatrics*. 1999;103(1 Suppl E):384-393.

62. Øvretveit J, Bate P, Cleary P, et al. Quality collaboratives: lessons from research. *Qual Saf Health Care*. 2002;11(4):345-351. doi:10.1136/qhc.11.4.345.

63. Schouten LMT, Hulscher MEJL, van Everdingen JJE, Huijsman R, Grol RPTM. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ*. 2008;336(7659):1491-1494. doi:10.1136/bmj.39570.749884.BE.

64. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. Understanding the Components of Quality Improvement Collaboratives: A Systematic Literature Review. *Milbank Quarterly*. 2013;91(2):354-394. doi:10.1111/milq.12016.

65. Parry GJ. Personal communication. September 2015.

66. Zanetti G, Giardina R, Platt R. Intraoperative redosing of cefazolin and risk for surgical site infection in cardiac surgery. *Emerging Infect Dis*. 2001;7(5):828-831. doi:10.3201/eid0705.017509.

67. Flamm BL, Berwick DM, Kabcenell A. Reducing cesarean section rates safely: lessons from a “breakthrough series” collaborative. *Birth*. 1998;25(2):117-124.

68. Kaplan HC, Brady PW, Dritz MC, et al. The influence of context on quality improvement success in health care: a systematic review of the literature. *Milbank*

Quarterly. 2010;88(4):500-559. doi:10.1111/j.1468-0009.2010.00611.x.

69. Berwick DM. Developing and testing changes in delivery of care. *Ann Intern Med*. 1998;128(8):651-656.

70. Hempel S, Maggard-Gibbons M, Nguyen DK, et al. Wrong-Site Surgery, Retained Surgical Items, and Surgical Fires : A Systematic Review of Surgical Never Events. *JAMA Surg*. 2015;150(8):796-805. doi:10.1001/jamasurg.2015.0301.

71. Milstein A, Gilbertson E. American medical home runs. *Health Aff (Millwood)*. 2009;28(5):1317-1326.

72. Nembhard IM. All teach, all learn, all improve?: the role of interorganizational learning in quality improvement collaboratives. *Health Care Manage Rev*. 2012;37(2):154-164. doi:10.1097/HMR.0b013e31822af831.

73. Homer CJ, Forbes P, Horvitz L, Peterson LE, Wypij D, Heinrich P. Impact of a quality improvement program on care and outcomes for children with asthma. *Arch Pediatr Adolesc Med*. 2005;159(5):464-469.

74. Horbar JD, Carpenter JH, Buzas J, et al. Collaborative quality improvement to promote evidence based surfactant for preterm infants: a cluster randomised trial. *BMJ*. 2004;329(7473):1004–0. doi:10.1136/bmj.329.7473.1004.

75. Baier RR, Gifford DR, Patry G, et al. Ameliorating pain in nursing homes: a collaborative quality-improvement project. *J Am Geriatr Soc*. 2004;52(12):1988-1995. doi:10.1111/j.1532-5415.2004.52553.x.

76. Pierce-Bulger M, Nighswander T. Nutaqsiivik--an approach to reducing infant mortality using quality improvement principles. *Quality Management in Healthcare*. 2001;9(3):40-46.

77. Horbar JD, Rogowski J, Plsek PE, et al. Collaborative quality improvement for neonatal intensive care. NIC/Q Project Investigators of the Vermont Oxford Network. *Pediatrics*. 2001;107(1):14-22.

78. Benedetti R, Flock B, Pedersen S, Ahern M. Improved clinical outcomes for fee-for-service physician practices participating in a diabetes care collaborative. *Joint Commission Journal on Quality and Safety*. 2004;30(4):187-194.

79. Schonlau M, Mangione-Smith R, Chan KS, et al. Evaluation of a quality improvement collaborative in asthma care: does it improve processes and outcomes of care? *Ann Fam Med*. 2005;3(3):200-208. doi:10.1370/afm.269.

80. Mangione-Smith R, Schonlau M, Chan KS, et al. Measuring the Effectiveness of a Collaborative for Quality Improvement in Pediatric Asthma Care: Does Implementing the Chronic Care Model Improve Processes and Outcomes of Care? *Ambulatory Pediatrics*. 2005;5(2):75. doi:10.1367/A04-106R.1.

81. Power M, Tyrrell PJ, Rudd AG, et al. Did a quality improvement collaborative make stroke care better? A cluster randomized trial. *Implementation Science*. 2014;9(1):1-9. doi:10.1186/1748-5908-9-40.

82. Rogowski JA, Horbar JD, Plsek PE, et al. Economic implications of neonatal intensive care unit collaborative quality improvement. *Pediatrics*. 2001;107(1):23-29.

83. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Niessen LW, Grol RP. Short- and long-term effects of a quality improvement collaborative on diabetes management. *Implementation Science*. 2010;5(1):94. doi:10.1186/1748-5908-5-94.

84. Schouten LMT, Niessen LW, van de Pas JWAM, Grol RPTM, Hulscher MEJL. Cost-Effectiveness of a Quality Improvement Collaborative Focusing on Patients With Diabetes. *Med Care*. 2010;48(10):884-891. doi:10.2307/25750574?ref=no-x-route:70089ef3cbcda172510ebbc5760a2d1c.

85. Barceló A, Cafiero E, de Boer M, et al. Using collaborative learning to improve diabetes care and outcomes: the VIDA project. *Prim Care Diabetes*. 2010;4(3):145-153. doi:10.1016/j.pcd.2010.04.005.

86. Margolis PA, McLearn KT, Earls MF, et al. Assisting primary care practices in using office systems to promote early childhood development. *Ambul Pediatr*. 2008;8(6):383-387. doi:10.1016/j.ambp.2008.06.007.

87. Asch S, Baker DW, Keesey JW, et al. Does the collaborative model improve care for chronic heart failure? *Med Care*. 2005;43(7):667-675.

88. Baker DW, Asch S, Keesey JW, et al. Differences in education, knowledge, self-management activities, and health outcomes for patients with heart failure cared for under the chronic disease model: the improving chronic illness care evaluation. *J Card Fail*. 2005;11(6):405-413. doi:10.1016/j.cardfail.2005.03.010.

89. Howard DH, Siminoff LA, McBride V, Lin M. Does quality improvement work? Evaluation of the Organ Donation Breakthrough Collaborative. - PubMed - NCBI. *Health Serv Res*. 2007;42(6p1):2160-2173. doi:10.1111/j.1475-6773.2007.00732.x.

90. Landon BE, Wilson IB, McInnes K, et al. Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV study. *Annals of Internal Medicine*. 2004;140(11):887-896.

91. O'Connor GT, Plume SK, Olmstead EM, et al. A regional prospective study of in-hospital mortality associated with coronary artery bypass grafting. The Northern New England Cardiovascular Disease Study Group. *JAMA*. 1991;266(6):803-809.

92. O'Connor GT, Plume SK, Olmstead EM, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA*. 1996;275(11):841-846.

93. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff (Millwood)*. 2002;21(4):40-51.

94. Horbar JD. The Vermont Oxford Network: Evidence-Based Quality Improvement for Neonatology. *Pediatrics*. 1999;103(Supplement E1):350-359. doi:10.1016/S0146-0005(05)80032-1.

95. Vermont Oxford Network. What Is Vermont Oxford Network? public.vtoxford.org. https://public.vtoxford.org/about-us/. Published 2015. Accessed October 22, 2015.

96. Payne NR, Finkelstein MJ, Liu M, Kaempf JW, Sharek PJ, Olsen S. NICU practices and outcomes associated with 9 years of quality improvement collaboratives. *Pediatrics*. 2010;125(3):437-446. doi:10.1542/peds.2009-1272.

97. Weiss KB, Mendoza G, Schall M, Berwick DM, Roessner J. *Improving Asthma Care in Children and Adults: Breakthrough Series Guide*. Boston: Institute for Healthcare Improvement; 1997.

98. Nolan TW, Schall MW, Berwick DM, Roessner J. *Reducing Delays and Waiting Times Throughout the Healthcare System: Breakthrough Series Guide*. Boston: Institute for Healthcare Improvement; 1996.

99. Leape LL, Kabcenell AI, Gandhi TK, Carver P, Nolan TW, Berwick DM. Reducing adverse drug events: lessons from a breakthrough series collaborative. *Jt Comm J Qual Improv*. 2000;26(6):321-331.

100. Nugent WC, Kilo CM, Ross CS, Marrin CA, Berwick DM, Roessner J. *Improving Outcomes and Reducing Costs in Adult Cardiac Surgery*. Boston: Institute for Healthcare Improvement; 1999:1-147.

101. Doran KA, Henry SA, Anderson BJ. Breakthrough change for adult cardiac surgery in a community-based cardiovascular program. *Quality Management in Healthcare*. 1998;6(4):29-36.

102. Rainey TG, Kabcenell A, Berwick DM, Roessner J. *Reducing Costs and Improving Outcomes in Adult Intensive Care: Breakthrough Series Guide*. Boston: Institute for Healthcare Improvement; 1998.

103. Besserman E, Brennan M, Brown PA3, Cleaves S, Nemeth WJ. Multidisciplinary achievement: the collaborative approach to rapid cycle ICU and hospital change. *Quality Management in Healthcare*. 1998;6(4):43-51.

104. Lynn J, Schall MW, Milne C, Nolan KM, Kabcenell A. Quality improvements in end of life care: insights from two collaboratives. *Jt Comm J Qual Improv*. 2000;26(5):254-267.

105. Sperl-Hillen J, O'Connor PJ, Carlson RR, et al. Improving diabetes care in a large health care system: an enhanced primary care approach. *Jt Comm J Qual Improv*. 2000;26(11):615-622.

106. Solberg LI. If you've seen one quality improvement collaborative ... *Ann Fam Med*. 2005;3(3):198-199. doi:10.1370/afm.304.

107. Sacks GD, Shannon EM, Dawes AJ, et al. Teamwork, communication and safety climate: a systematic review of interventions to improve surgical culture. *BMJ Quality & Safety*. 2015;24(7):458-467. doi:10.1136/bmjqs-2014-003764.

108. Hulscher MEJL, Schouten LMT, Grol RPTM, Buchan H. Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ Quality & Safety*. 2013;22(1):19-31. doi:10.1136/bmjqs-2011-000651.

109. Wilson T, Berwick DM, Cleary PD. What do collaborative improvement projects do? Experience from seven countries. *Joint Commission Journal on Quality and Safety*. 2003.

110. Maggard-Gibbons M. The use of report cards and outcome measurements to improve the safety of surgical care: the American College of Surgeons National Surgical Quality Improvement Program. *BMJ Quality & Safety*. 2014;23(7):589-599. doi:10.1136/bmjqs-2013-002223.

111. Malloy E. Personal communication. May 2015.

112. Gillissen F, Hoff C, Maessen JMC, et al. Structured synchronous implementation of an enhanced recovery program in elective colonic surgery in 33 hospitals in The Netherlands. *World J Surg*. 2013;37(5):1082-1093. doi:10.1007/s00268-013-1938-4.

113. van Veen-Berkx E, Bitter J, Kazemier G, Scheffer GJ, Gooszen HG. Multidisciplinary teamwork improves use of the operating room: a multicenter study. - PubMed - NCBI. *J Am Coll Surg*. 2015;220(6):1070-1076. doi:10.1016/j.jamcollsurg.2015.02.012.

114. Fung-Kee-Fung M, Watters J, Crossley C, et al. Regional collaborations as a tool for quality improvements in surgery: a systematic review of the literature. *Ann Surg*. 2009;249(4):565-572. doi:10.1097/SLA.0b013e31819ec608.

115. Dellinger EP, Hausmann SM, Bratzler DW, et al. Hospitals collaborate to decrease surgical site infections. *Am J Surg*. 2005;190(1):9-15. doi:10.1016/j.amjsurg.2004.12.001.

116. Kritchevsky SB, Braun BI, Bush AJ, et al. The effect of a quality improvement collaborative to improve antimicrobial prophylaxis in surgical patients: a randomized trial. *Annals of Internal Medicine*. 2008;149(7):472–80–W89–93.

117. Campbell DA, Englesbe MJ, Kubus JJ, et al. Accelerating the pace of surgical quality improvement: the power of hospital collaboration. *Arch Surg*. 2010;145(10):985-991. doi:10.1001/archsurg.2010.220.

118. Share DA, Campbell DA, Birkmeyer N, et al. How a regional collaborative of hospitals and physicians in Michigan cut costs and improved the quality of care. *Health Aff*

(Millwood). 2011;30(4):636-645. doi:10.1377/hlthaff.2010.0526.

119. Tavris DR, Dahl J, Gordon D, et al. Evaluation of a local cooperative project to improve postoperative pain management in Wisconsin hospitals. *Quality Management in Healthcare*. 1999;7(2):20-27.

120. Birkmeyer NJO, Share D, Campbell DA, Prager RL, Moscucci M, Birkmeyer JD. Partnering with payers to improve surgical quality: the Michigan plan. *Surgery*. 2005;138(5):815-820. doi:10.1016/j.surg.2005.06.037.

121. Campbell DA Jr, Kubus JJ, Henke PK, Hutton M, Englesbe MJ. The Michigan Surgical Quality Collaborative: a legacy of Shukri Khuri. *The American Journal of Surgery*. 2009;198(5):S49-S55. doi:10.1016/j.amjsurg.2009.08.002.

122. Campbell DA, Dellinger EP. Multihospital collaborations for surgical quality improvement. *JAMA: The Journal of the American Medical Association*. 2009;302(14):1584-1585. doi:10.1001/jama.2009.1474.

123. Campbell DA, Englesbe MJ. How can the American College of Surgeons-National Surgical Quality Improvement Program help or hinder the general surgeon? *Adv Surg*. 2008;42:169-181.

124. Englesbe MJ, Brooks L, Kubus J, et al. A statewide assessment of surgical site infection following colectomy: the role of oral antibiotics. *Ann Surg*. 2010;252(3):514–9–discussion519–20. doi:10.1097/SLA.0b013e3181f244f8.

125. Krell RW, Girotti ME, Fritze D, Campbell DA, Hendren S. Hospital readmissions after colectomy: a population-based study. *J Am Coll Surg*. 2013;217(6):1070-1079. doi:10.1016/j.jamcollsurg.2013.07.403.

126. Hendren S, Fritze D, Banerjee M, et al. Antibiotic choice is independently associated with risk of surgical site infection after colectomy: a population-based cohort study. *Ann Surg*. 2013;257(3):469-475. doi:10.1097/SLA.0b013e31826c4009.

127. Kim EK, Sheetz KH, Bonn J, et al. A statewide colectomy experience: the role of full bowel preparation in preventing surgical site infection. *Ann Surg*. 2014;259(2):310-314. doi:10.1097/SLA.0b013e3182a62643.

128. Hendren S, Englesbe MJ, Brooks L, Kubus J, Yin H, Campbell DAJ. Prophylactic antibiotic practices for colectomy in Michigan. *Am J Surg*. 2011;201(3):290–3–discussion293–4. doi:10.1016/j.amjsurg.2010.08.024.

129. Waits SA, Reames BN, Sheetz KH, Englesbe MJ, Campbell DA. Anticipating the Effects of Medicaid Expansion on Surgical Care. *JAMA Surg*. 2014;149(7):–. doi:10.1001/jamasurg.2014.222.

130. Reames BN, Krell RW, Campbell DAJ, Dimick JB. A checklist-based intervention to improve surgical outcomes in Michigan: evaluation of the Keystone Surgery program.

JAMA Surg. 2015;150(3):208-215. doi:10.1001/jamasurg.2014.2873.

131. Henke PK, Kubus J, Englesbe MJ, Harbaugh C, Campbell DA. A statewide consortium of surgical care: a longitudinal investigation of vascular operative procedures at 16 hospitals. *Surgery*. 2010;148(4):883–89–discussion889–92. doi:10.1016/j.surg.2010.07.009.

132. Healy MA, Krell RW, Abdelsattar ZM, et al. Pancreatic Resection Results in a Statewide Surgical Collaborative. *Ann Surg Oncol*. 2015;22(8):2468-2474. doi:10.1245/s10434-015-4529-9.

133. Flum DR, Fisher N, Thompson J, Marcus-Smith M, Florence M, Pellegrini CA. Washington State's approach to variability in surgical processes/Outcomes: Surgical Clinical Outcomes Assessment Program (SCOAP). *Surgery*. 2005;138(5):821-828. doi:10.1016/j.surg.2005.07.026.

134. Kwon S, Florence M, Grigas P, et al. Creating a learning healthcare system in surgery: Washington State's Surgical Care and Outcomes Assessment Program (SCOAP) at 5 years. *Surgery*. 2012;151(2):146-152. doi:10.1016/j.surg.2011.08.015.

135. Flum D. Personal communication. April 2016.

136. Cuschieri J, Florence M, Flum DR, et al. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. *Ann Surg*. 2008;248(4):557-563. doi:10.1097/SLA.0b013e318187aeca.

137. Drake FT, Florence MG, Johnson MG, et al. Progress in the diagnosis of appendicitis: a report from Washington State's Surgical Care and Outcomes Assessment Program. *Ann Surg*. 2012;256(4):586-594. doi:10.1097/SLA.0b013e31826a9602.

138. Drake FT, Alfonso R, Bhargava P, et al. Enteral contrast in the computed tomography diagnosis of appendicitis: comparative effectiveness in a prospective surgical cohort. *Ann Surg*. 2014;260(2):311-316. doi:10.1097/SLA.0000000000000272.

139. Kwon S, Morris A, Billingham R, et al. Routine leak testing in colorectal surgery in the Surgical Care and Outcomes Assessment Program. *Arch Surg*. 2012;147(4):345-351. doi:10.1001/archsurg.2012.12.

140. Hakkarainen TW, Steele SR, Bastaworous A, et al. Nonsteroidal Anti-inflammatory Drugs and the Risk for Anastomotic Failure. *JAMA Surg*. 2015;150(3):223-228. doi:10.1001/jamasurg.2014.2239.

141. Kwon S, Thompson R, Florence M, et al. β-blocker continuation after noncardiac surgery: a report from the surgical care and outcomes assessment program. *Arch Surg*. 2012;147(5):467-473. doi:10.1001/archsurg.2011.1698.

142. Surgical Care and Outcomes Assessment Program (SCOAP) Collaborative, Kwon S, Billingham R, et al. Adoption of laparoscopy for elective colorectal resection: a report

from the Surgical Care and Outcomes Assessment Program. *J Am Coll Surg*. 2012;214(6):909–18.e1. doi:10.1016/j.jamcollsurg.2012.03.010.

143. Simianu VV, Bastawrous AL, Billingham RP, et al. Addressing the appropriateness of elective colon resection for diverticulitis: a report from the SCOAP CERTAIN collaborative. *Ann Surg*. 2014;260(3):533–8–discussion538–9. doi:10.1097/SLA.0000000000000894.

144. Kwon S, Meissner M, Symons R, et al. Perioperative pharmacologic prophylaxis for venous thromboembolism in colorectal surgery. - PubMed - NCBI. *J Am Coll Surg*. 2011;213(5):596–603.e1. doi:10.1016/j.jamcollsurg.2011.07.015.

145. Colorectal Writing Group for Surgical Care and Outcomes Assessment Program–Comparative Effectiveness Research Translation Network (SCOAP-CERTAIN) Collaborative, Nelson DW, Simianu VV, et al. Thromboembolic Complications and Prophylaxis Patterns in Colorectal Surgery. *JAMA Surg*. 2015;150(8):712-720. doi:10.1001/jamasurg.2015.1057.

146. Breslin TM, Caughran J, Pettinga J, et al. Improving breast cancer care through a regional quality collaborative. *Surgery*. 2011;150(4):635-642. doi:10.1016/j.surg.2011.07.071.

147. Rickles AS, Iannuzzi JC, Kelly KN, et al. Anastomotic leak or organ space surgical site infection: What are we missing in our quality improvement programs? *Surgery*. 2013;154(4):680-689. doi:10.1016/j.surg.2013.06.035.

148. Upstate New York Surgical Quality Initiative. The Power of Collaboration. unysqi.org. http://www.unysqi.org/home.aspx. Published 2015. Accessed August 6, 2015.

149. Cofer JB. The Tennessee Surgical Quality Collaborative: a model for improving quality of care and reducing health care costs. *Tenn Med*. 2012;105(8):31-33.

150. Guillamondegui OD, Gunter OL, Hines L, et al. Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to improve surgical outcomes. *J Am Coll Surg*. 2012;214(4):709–14–discussion714–6. doi:10.1016/j.jamcollsurg.2011.12.012.

151. Daley BJ, Cecil W, Clarke PC, Cofer JB, Guillamondegui OD. How slow is too slow? Correlation of operative time to complications: an analysis from the Tennessee Surgical Quality Collaborative. *J Am Coll Surg*. 2015;220(4):550-558. doi:10.1016/j.jamcollsurg.2014.12.040.

152. Florida Hospital Association. The Florida Surgical Care Initiative: A Closer Look - May 2010. www.fha.org. http://www.fha.org/showDocument.aspx?f=FSCI_A_Closer_Look_013113.pdf. Published May 2010. Accessed August 6, 2015.

153. Florida Surgical Care Initiative. www.fha.org.

http://www.fha.org/showDocument.aspx?f=FSCI-Brochure-013113.pdf. Accessed August 6, 2015.

154. Florida Hospital Association. Five Years of Quality. http://www.fha.org/modules/1/showReport.aspx?dId=1&rId=89. Published 2013. Accessed August 6, 2015.

155. Florida Hospital Association. Florida Surgical Care Initiative. www.fha.org. http://www.fha.org/health-care-issues/quality-and-safety/florida-surgical-care-initiative-.aspx. Accessed August 6, 2015.

156. Tepas JJ, Kerwin AJ, deVilla J, Nussbaum MS. Macro vs micro level surgical quality improvement: a regional collaborative demonstrates the case for a national NSQIP initiative. *J Am Coll Surg*. 2014;218(4):599-604. doi:10.1016/j.jamcollsurg.2013.12.017.

157. Bradley SM, Bell M, Ahuja V. Multifaceted Differences in Implementation of Practices for Prevention of Colorectal and Bariatric Surgical Site Infections. *Pa Patient Saf Advis*. 2012;9(4):136-142.

158. Davis J, Catanzaro M, Edelmayer C, Martin L. Surgical Site Infection Prevention Utilizing Patient Screening and Decolonization: The PA-HEN SSI Prevention Collaboration. *Pa Patient Saf Advis*. 2014;11(3):131-135.

159. Marsden J, van Dijk M, Doris P, Krause C, Cochrane D. Improving care for British Columbians: the critical role of physician engagement. *Healthc Q*. 2012;15 Spec No:51-55.

160. Surgical Quality Action Network. The National Surgical Quality Improvement Program in BC: 2.5 Years After Start-Up. bcpsqc.ca. https://bcpsqc.ca//documents/2012/10/NSQIP-2.5-years-later_January-2014_final_v3.pdf. Published January 2014. Accessed August 6, 2015.

161. Surgical Quality Action Network. Tackling Complex Problems with Team-Based Solutions: NSQIP in BC 2014. bcpsqc.ca. https://bcpsqc.ca//documents/2012/10/Tackling-Complex-Problems-with-Team-Based-Solutions.pdf. Accessed August 6, 2015.

162. British Columbia Patient Safety Quality Council. Surgical Quality Action Network. bcpsqc.ca. https://bcpsqc.ca/clinical-improvement/sqan/. Accessed August 6, 2015.

163. Connecticut Chapter of the American College of Surgeons Profession Association, Inc.,. CtSQC. http://ctacs.org/?page_id=29. Published March 8, 2012. Accessed August 6, 2015.

164. Corvo PR. We Are Building a Cathedral. In:; 2015.

165. Corvo PR, Tasik C. Surgical Quality Initiatives in Connecticut: We Are Building a Cathedral. *Conn Med*. 2015;79(6):361-363.

166. Wilson B. Colorectal Pathway: A Template for the Georgia Surgical Quality Collaborative. In:; 2015.

167. Illinois Surgical Quality Improvement Collaborative. isqic.org. http://www.isqic.org/. Published 2015. Accessed August 5, 2015.

168. Bilimoria KY. Illinois Surgical Quality Improvement Collaborative. In:; 2015.

169. Birkmeyer NJO, Birkmeyer JD. Strategies for improving surgical quality--should payers reward excellence or effort? *N Engl J Med*. 2006;354(8):864-870. doi:10.1056/NEJMsb053364.

170. Englesbe MJ, Dimick JB, Sonnenday CJ, Share DA, Campbell DA. The Michigan surgical quality collaborative: will a statewide quality improvement initiative pay for itself? *Ann Surg*. 2007;246(6):1100-1103. doi:10.1097/SLA.0b013e31815c3fe5.

171. Smith M, Hussain A, Xiao J, et al. The Importance of Improving the Quality of Emergency Surgery for a Regional Quality Collaborative. *Ann Surg*. 2013;257(4):596-602. doi:10.1097/SLA.0b013e3182863750.

172. Moscucci M, Rogers EK, Montoye C, et al. Association of a continuous quality improvement initiative with practice and outcome variations of contemporary percutaneous coronary interventions. *Circulation*. 2006;113(6):814-822. doi:10.1161/CIRCULATIONAHA.105.541995.

173. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med*. 2006;355(26):2725-2732. doi:10.1056/NEJMoa061115.

174. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood)*. 2005;24(1):138-150. doi:10.1377/hlthaff.24.1.138.

175. Schuster MA, McGlynn EA, Brook RH. How Good Is the Quality of Health Care in the United States? *Milbank Q*. 2005;83(4):843-895. doi:10.2307/30045644?ref=no-x-route:286a9a61c37aa6a643a56bf1708ee91a.

176. Jencks SF, Cuerdon T, Burwen DR, et al. Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile at State and National Levels. *JAMA*. 2000;284(13):1670-1676. doi:10.1001/jama.284.13.1670.

177. Schoen C, Davis K, How SKH, Schoenbaum SC. U.S. health system performance: a national scorecard. *Health Affairs*. 2006;25(6):w457-w475. doi:10.1377/hlthaff.25.w457.

178. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med*. 1988;109(7):582-589.

179. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' Perception of Hospital Care in the

United States. *N Engl J Med*. 2008;359(18):1921-1931. doi:10.1056/NEJMsa0804116.

180. Institute of Medicine (US) Committee on Quality of Health Care in America. *Crossing the Quality Chasm: a New Health System for the 21st Century*. Washington (DC): National Academies Press (US); 2001. doi:10.17226/10027.

181. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. In: Vol 280. 1998:1000-1005.

182. Advisory Commission on Consumer Protection and Quality in the Health Care Industry. Quality First: Better Health Care for All Americans. archive.ahrq.gov. http://archive.ahrq.gov/hcqual/final/append_a.html. Published July 17, 1998. Accessed February 20, 2016.

183. Chassin MR, Loeb JM. The ongoing quality improvement journey: next stop, high reliability. *Health Affairs*. 2011;30(4):559-568. doi:10.1377/hlthaff.2011.0076.

184. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med*. 2005;61(2):417-430. doi:10.1016/j.socscimed.2004.12.001.

185. Sofaer S. Qualitative methods: what are they and why use them? *Health Serv Res*. 1999;34(5 Pt 2):1101-1118.

186. Ryan GW, Bernard HR. Techniques to Identify Themes. *Field Methods*. 2003;15(1):85-109.

187. Mills PD, Weeks WB. Characteristics of successful quality improvement teams: lessons from five collaborative projects in the VHA. *Joint Commission Journal on Quality and Safety*. 2004;30(3):152-162.

188. Ayers LR, Beyea SC, Godfrey MM, Harper DC, Nelson EC, Batalden PB. Quality improvement learning collaboratives. *Quality Management in Healthcare*. 2005;14(4):234-247.

189. Dückers MLA, Wagner C, Groenewegen PP. Developing and testing an instrument to measure the presence of conditions for successful implementation of quality improvement collaboratives. *BMC Health Serv Res*. 2008;8(1):172. doi:10.1186/1472-6963-8-172.

190. Dückers MLA, Spreeuwenberg P, Wagner C, Groenewegen PP. Exploring the black box of quality improvement collaboratives: modelling relations between conditions, applied changes and outcomes. *Implementation Science*. 2009;4(1):74. doi:10.1186/1748-5908-4-74.

191. Campbell DA. Quality improvement is local. *J Am Coll Surg*. 2009;209(1):141-143. doi:10.1016/j.jamcollsurg.2009.03.012.

192. Nembhard IM. Learning and improving in quality improvement collaboratives: which collaborative features do participants value most? *Health Serv Res*. 2009;44(2 Pt 1):359-378. doi:10.1111/j.1475-6773.2008.00923.x.

193. Schouten LM, Grol RP, Hulscher ME. Factors influencing success in quality-improvement collaboratives: development and psychometric testing of an instrument. *Implementation Science*. 2010;5(1):84. doi:10.1186/1748-5908-5-84.

194. Øvretveit J. How to run an effective improvement collaborative. *Int J Health Care Qual Assur*. 2013;15(5):192-196. doi:10.1108/09526860210437403.

195. Kasper JF, Plume SK, O'Connor GT. A methodology for QI in the coronary artery bypass grafting procedure involving comparative process analysis. *QRB Qual Rev Bull*. 1992;18(4):129-133.

196. Perkins DN, Salomon G. Teaching for Transfer. *Educational Leadership*. 1988;46(1):22.

197. Perkins DN, Salomon G. The science and art of transfer. In: Costa AL, Bellanca J, Fogarty R, eds. *If Minds Matter: a Foreword to the Future*. Cheltenham, Victoria: Hawker Brownlow Education; 1992:201-210.

198. Barker JA. *Future Edge: Discovering the New Paradigms of Success*. New York, NY: William Morrow; 1992.

199. Weiner BJ, Amick H, Lee S-YD. Conceptualization and measurement of organizational readiness for change: a review of the literature in health services research and other fields. *Med Care Res Rev*. 2008;65(4):379-436. doi:10.1177/1077558708317802.

200. Weiner BJ. A theory of organizational readiness for change. *Implementation Science*. 2009;4(1):67. doi:10.1186/1748-5908-4-67.

201. Scaccia JP, Cook BS, Lamont A, et al. A practical implementation science heuristic for organizational readiness: R = MC. *J Community Psychol*. 2015;43(4):484-501. doi:10.1002/jcop.21698.

202. Mangione-Smith R. Personal communication. March 2016.

203. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Quality & Safety*. 2012;21(1):13-20. doi:10.1136/bmjqs-2011-000010.

204. Rycroft-Malone J, Kitson A, Harvey G, et al. Ingredients for change: revisiting a conceptual framework. *Qual Saf Health Care*. 2002;11(2):174-180.

205. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*.

2009;4(1):50. doi:10.1186/1748-5908-4-50.

206. Fisher ES, Shortell SM, Savitz LA. Implementation Science: A Potential Catalyst for Delivery System Reform. *JAMA*. 2016;315(4):339-340. doi:10.1001/jama.2015.17949.

207. Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. Reality check for checklists. *Lancet*. August 8, 2009:444-445.

208. Taylor SL, Dy S, Foy R, et al. What context features might be important determinants of the effectiveness of patient safety practice interventions? *BMJ Quality & Safety*. 2011;20(7):611-617. doi:10.1136/bmjqs.2010.049379.

209. Alexander JA, Weiner BJ, Shortell SM, Baker LC, Becker MP. The Role of Organizational Infrastructure in Implementation of Hospitals' Quality Improvement. *Hospital Topics*. 2010;84(1):11-21. doi:10.3200/HTPS.84.1.11-21.

210. 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(2):281-315. doi:10.1111/1468-0009.00206.

211. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*. 2004;82(4):581-629. doi:10.1111/j.0887-378X.2004.00325.x.

212. Helfrich CD, Li Y-F, Sharp ND, Sales AE. Organizational readiness to change assessment (ORCA): development of an instrument based on the Promoting Action on Research in Health Services (PARIHS) framework. *Implementation Science*. 2009;4(1):38. doi:10.1186/1748-5908-4-38.

213. Shea CM, Jacobs SR, Esserman DA, Bruce K, Weiner BJ. Organizational readiness for implementing change: a psychometric assessment of a new measure. *Implementation Science*. 2014;9(1):1. doi:10.1186/1748-5908-9-7.

214. Gagnon M-P, Attieh R, Ghandour EK, et al. A systematic review of instruments to assess organizational readiness for knowledge translation in health care. Jeyaseelan K, ed. *PLoS ONE*. 2014;9(12):e114338. doi:10.1371/journal.pone.0114338.

215. American College of Surgeons Trauma Quality Improvement Program. *ACS TQIP Aggregate Report: Fall 2015*. Chicago, IL; 2015:1-44.

216. Merkow RP, Hall BL, Cohen ME, et al. Relevance of the c-statistic when evaluating risk-adjustment models in surgery. - PubMed - NCBI. *J Am Coll Surg*. 2012;214(5):822-830.

217. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351:h3868. doi:10.1136/bmj.h3868.

218. Dawes AJ, Sacks GD, Cryer HG, et al. Compliance With Evidence-Based Guidelines

and Interhospital Variation in Mortality for Patients With Severe Traumatic Brain Injury. *JAMA Surg*. 2015;150(10):965-972. doi:10.1001/jamasurg.2015.1678.

219. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Annals of Internal Medicine*. 2015;162(10):735-736. doi:10.7326/L15-5093-2.

220. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996;58(1):267-288. doi:10.2307/2346178?ref=search-gateway:24e50819b6a5cf155f1c6f8a864c660b.

221. Adams JL. *The Reliability of Provider Profiling: a Tutorial*. Santa Monica, CA: RAND Corporation; 2009.

222. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res*. 2010;45(6 Pt 1):1614-1629. doi:10.1111/j.1475-6773.2010.01158.x.

223. Haider AH, Hashmi ZG, Zafar SN, et al. Developing best practices to study trauma outcomes in large databases: an evidence-based approach to determine the best mortality risk adjustment model. *J Trauma Acute Care Surg*. 2014;76(4):1061-1069. doi:10.1097/TA.0000000000000182.

224. Osborne NH, Ko CY, Upchurch GR Jr, Dimick JB. Evaluating parsimonious risk-adjustment models for comparing hospital outcomes with vascular surgery. *J Vasc Surg*. 2010;52(2):400-405. doi:10.1016/j.jvs.2010.02.293.

225. Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg*. 2010;210(4):503-508. doi:10.1016/j.jamcollsurg.2010.01.018.

226. Shafi S, Ahn C, Parks J, et al. Quality of Care Within a Trauma Center Is not Altered by Injury Type. *J Trauma*. 2010;68(3):716-720. doi:10.1097/TA.0b013e3181a7bec0.

227. Shwartz M, Cohen AB, Restuccia JD, et al. How well can we identify the high-performing hospital? *Med Care Res Rev*. 2011;68(3):290-310. doi:10.1177/1077558710386115.

228. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP Modeling for Evaluation of Surgical Quality and Risk: Patient Risk Adjustment, Procedure Mix Adjustment, Shrinkage Adjustment, and Surgical Focus. *J Am Coll Surg*. 2013;217(2):336–346.e1. doi:10.1016/j.jamcollsurg.2013.02.027.

229. Livingston E, Cao J, Dimick JB. Tread Carefully With Stepwise Regression. *Arch Surg*. 2010;145(11):1039-1040. doi:10.1001/archsurg.2010.240.

230. Machado-Aranda DA, Jakubus JL, Wahl WL, et al. Reduction in Venous Thromboembolism Events: Trauma Performance Improvement and Loop Closure

Through Participation in a State-Wide Quality Collaborative. *J Am Coll Surg*. 2015;221(3):661-668. doi:10.1016/j.jamcollsurg.2015.05.006.

231. Shafi S, Nathens AB, Cryer HG, et al. The Trauma Quality Improvement Program of the American College of Surgeons Committee on Trauma. *J Am Coll Surg*. 2009;209(4):521-530.e521. doi:10.1016/j.jamcollsurg.2009.07.001.

232. Selassie AW, Zaloshnja E, Langlois JA, Miller T, Jones P, Steiner C. Incidence of long-term disability following traumatic brain injury hospitalization, United States, 2003. *J Head Trauma Rehabil*. 2008;23(2):123-131. doi:10.1097/01.HTR.0000314531.30401.39.

233. Farhad K, Khan HMR, Ji AB, Yacoub HA, Qureshi AI, Souayah N. Trends in Outcomes and Hospitalization Costs for Traumatic Brain Injury in Adult Patients in the United States. *Journal of Neurotrauma*. 2013;30(2):84-90. doi:10.1089/neu.2011.2283.

234. Ghajar J, Hariri RJ, Narayan RK, Iacono LA, Firlik K, Patterson RH. Survey of critical care management of comatose, head-injured patients in the United States. *Crit Care Med*. 1995;23(3):560-567.

235. Brain Trauma Foundation, American Association of Neurological Surgeons, Congress of Neurological Surgeons. Guidelines for the management of severe traumatic brain injury. *Journal of Neurotrauma*. 2007;24 Suppl 1(supplement 1):S1-S106. doi:10.1089/neu.2007.9999.

236. Bratton SL, Chestnut RM, Ghajar J, et al. VI. Indications for Intracranial Pressure Monitoring. *Journal of Neurotrauma*. 2007;24(supplement 1):S–37–S–44. doi:10.1089/neu.2007.9990.

237. Bullock MR, Chesnut R, Ghajar J, et al. Guidelines for the Surgical Management of Traumatic Brain Injury. *Neurosurgery*. 2006;58(3):S2–1–S2–112.

238. Fakhry SM, Trask AL, Waller MA, Watts DD, IRTC Neurotrauma Task Force. Management of brain-injured patients by an evidence-based medicine protocol improves outcomes and decreases hospital charges. *J Trauma*. 2004;56(3):492–9–discussion499–500.

239. Arabi YM, Haddad S, Tamim HM, et al. Mortality reduction after implementing a clinical practice guidelines–based management protocol for severe traumatic brain injury. *J Crit Care*. 2010;25(2):190-195. doi:10.1016/j.jcrc.2009.05.004.

240. Shafi S, Barnes SA, Rayan N, et al. Compliance with Recommended Care at Trauma Centers: Association with Patient Outcomes. *J Am Coll Surg*. 2014;219(2):189-198. doi:10.1016/j.jamcollsurg.2014.04.005.

241. Hesdorffer DC, Ghajar J. Marked improvement in adherence to traumatic brain injury guidelines in United States trauma centers. *J Trauma*. 2007;63(4):841–7–discussion847–8. doi:10.1097/TA.0b013e318123fc21.

242. Morse RB, Hall M, Fieldston ES, et al. Hospital-level compliance with asthma care quality measures at children's hospitals and subsequent asthma-related outcomes. *JAMA: The Journal of the American Medical Association*. 2011;306(13):1454-1460. doi:10.1001/jama.2011.1385.

243. Howell EA, Zeitlin J, Hebert PL, Balbierz A, Egorova N. Association between hospital-level obstetric quality indicators and maternal and neonatal morbidity. *JAMA: The Journal of the American Medical Association*. 2014;312(15):1531-1541. doi:10.1001/jama.2014.13381.

244. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA: The Journal of the American Medical Association*. 2006;296(22):2694-2702. doi:10.1001/jama.296.22.2694.

245. Ryan AM, suffix JFBSJ, Tompkins CP, Wallack SS. The Relationship between Medicare's Process of Care Quality Measures and Mortality. *Inquiry*. 2009;46(3):274-290. doi:10.2307/29773429?ref=no-x-route:bfe4bca0f1b8143eaf99da79dfeb7c61.

246. Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. *JAMA: The Journal of the American Medical Association*. 2010;303(24):2479-2485. doi:10.1001/jama.2010.841.

247. Bilimoria KY, Chung J, Ju MH, et al. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. *JAMA: The Journal of the American Medical Association*. 2013;310(14):1482-1489. doi:10.1001/jama.2013.280048.

248. Kernisan LP, Lee SJ, Boscardin WJ, Landefeld CS, Dudley RA. Association between hospital-reported Leapfrog Safe Practices Scores and inpatient mortality. *JAMA: The Journal of the American Medical Association*. 2009;301(13):1341-1348. doi:10.1001/jama.2009.422.

249. Qian F, Lustik SJ, Diachun CA, Wissler RN, Zollo RA, Glance LG. Association Between Leapfrog Safe Practices Score and Hospital Mortality in Major Surgery. *Med Care*. 2011;49(12):1082-1088. doi:10.2307/23053854?ref=no-x-route:bfe4bca0f1b8143eaf99da79dfeb7c61.

250. Glance LG, Dick AW, Osler TM, et al. Relationship between Leapfrog Safe Practices Survey and outcomes in trauma. *Arch Surg*. 2011;146(10):1170-1177. doi:10.1001/archsurg.2011.247.

251. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255(4):703-707. doi:10.1097/SLA.0b013e31824b46ff.

252. Silber JH, Rosenbaum PR, Brachet TJ, et al. The Hospital Compare Mortality Model and the Volume–Outcome Relationship. *Health Serv Res*. 2010;45(5 Pt 1):1148-1167. doi:10.1111/j.1475-6773.2010.01130.x.

253. Alali AS, Fowler RA, Mainprize TG, et al. Intracranial pressure monitoring in severe traumatic brain injury: results from the American College of Surgeons Trauma Quality Improvement Program. *Journal of Neurotrauma*. 2013;30(20):1737-1746. doi:10.1089/neu.2012.2802.

254. Biersteker HAR, Andriessen TMJC, Horn J, et al. Factors influencing intracranial pressure monitoring guideline compliance and outcome after severe traumatic brain injury. *Crit Care Med*. 2012;40(6):1914-1922. doi:10.1097/CCM.0b013e3182474bde.

255. Farahvar A, Gerber LM, Chiu Y-L, Carney N, Hartl R, Ghajar J. Increased mortality in patients with severe traumatic brain injury treated without intracranial pressure monitoring. *J Neurosurg*. 2012;117(4):729-734. doi:10.3171/2012.7.JNS111816.

256. Talving P, Karamanos E, Teixeira PG, et al. Intracranial pressure monitoring in severe head injury: compliance with Brain Trauma Foundation guidelines and effect on outcomes: a prospective study. *J Neurosurg*. 2013;119(5):1248-1254. doi:10.3171/2013.7.JNS122255.

257. Tang A, Pandit V, Fennell V, et al. Intracranial pressure monitor in patients with traumatic brain injury. *J Surg Res*. November 2014. doi:10.1016/j.jss.2014.11.017.

258. Schreiber MA, Aoki N, Scott BG, Beck JR. Determinants of mortality in patients with severe blunt head injury. *Arch Surg*. 2002;137(3):285-290.

259. Farahvar A, Gerber LM, Chiu Y-L, et al. Response to intracranial hypertension treatment as a predictor of death in patients with severe traumatic brain injury. *J Neurosurg*. 2011;114(5):1471-1478. doi:10.3171/2010.11.JNS101116.

260. Karamanos E, Teixeira PG, Sivrikoz E, et al. Intracranial pressure versus cerebral perfusion pressure as a marker of outcomes in severe head injury: a prospective evaluation. *Am J Surg*. 2014;208(3):363-371. doi:10.1016/j.amjsurg.2013.10.026.

261. Chesnut RM, Temkin N, Carney N, et al. A trial of intracranial-pressure monitoring in traumatic brain injury. *N Engl J Med*. 2012;367(26):2471-2481. doi:10.1056/NEJMoa1207363.

262. Khuri SF, Najjar SF, Daley J, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. *Ann Surg*. 2001;234(3):370–82–discussion382–3.

263. Ayanian JZ, Weissman JS. Teaching Hospitals and Quality of Care: A Review of the Literature. *Milbank Q*. 2002;80(3):569-593. doi:10.2307/3350565?ref=no-x-route:ae37c3f51bafc28d04d224227ec44a75.

264. Donabedian A. Evaluating the Quality of Medical Care. *The Milbank Memorial Fund Quarterly*. 1966;44(3):166-203. doi:10.1111/j.1468-0009.2005.00397.x.

265. Birkmeyer JD, Dimick JB, Birkmeyer NJO. Measuring the quality of surgical care:

structure, process, or outcomes?1 1No competing interests declared. *J Am Coll Surg*. 2004;198(4):626-632. doi:10.1016/j.jamcollsurg.2003.11.017.

266. Hawn MT, Itani KM, Gray SH, Vick CC, Henderson W, Houston TK. Association of Timely Administration of Prophylactic Antibiotics for Major Surgical Procedures and Surgical Site Infection. *J Am Coll Surg*. 2008;206(5):814-819. doi:10.1016/j.jamcollsurg.2007.12.013.

267. Berenholtz SM, Pronovost PJ, Lipsett PA, et al. Eliminating catheter-related bloodstream infections in the intensive care unit. *Crit Care Med*. 2004;32(10):2014-2020.

268. Berwick DM. Measuring surgical outcomes for improvement: was Codman wrong? *JAMA: The Journal of the American Medical Association*. 2015;313(5):469-470. doi:10.1001/jama.2015.4.

269. Chassin MR. Quality of care. Time to act. *JAMA*. 1991;266(24):3472-3473.

270. Davis K, Stremikis K, Squires D, Schoen C. *Mirror, Mirror on the Wall: How the Performance of the U.S. Health Care System Compares Internationally*. New York, NY: The Commonwealth Fund; 2014.

271. Tandon A, Murray CJL, Lauer JA, Evans DB. *Measuring Overall Health System Performance for 191 Countries*. Geneva: World Health Organization; 2000.

272. Cretin S, Shortell SM, Keeler EB. An evaluation of collaborative interventions to improve chronic illness care. Framework and study design. *Eval Rev*. 2004;28(1):28-51. doi:10.1177/0193841X03256298.

273. Bevan H, Penny J, Layton A. A "Breakthrough" approach to reducing delays and patient waiting times. *Clinician in Management*. 2000;9:27-31.

274. Nabozny MJ, Kruser JM, Steffens NM, et al. Constructing High-stakes Surgical Decisions: It's Better to Die Trying. *Ann Surg*. 2016;263(1):64-70. doi:10.1097/SLA.0000000000001081.

275. Kruser JM, Pecanac KE, Brasel KJ, et al. "And I think that we can fix it": mental models used in high-risk surgical decision making. *Ann Surg*. 2015;261(4):678-684. doi:10.1097/SLA.0000000000000714.

276. Malenka DJ, O'Connor GT. The Northern New England Cardiovascular Disease Study Group: a regional collaborative effort for continuous quality improvement in cardiovascular disease. *Jt Comm J Qual Improv*. 1998;24(10):594-600.

277. Campbell DA Jr. Personal communication. March 2016.

278. Batalden PB, Davidoff F. What is "quality improvement" and how can it transform healthcare? *Qual Saf Health Care*. 2007;16(1):2-3. doi:10.1136/qshc.2006.022046.

279. 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. doi:10.1056/NEJMsb070738.

280. Fan E, Laupacis A, Pronovost PJ, Guyatt GH, Needham DM. How to use an article about quality improvement. *JAMA: The Journal of the American Medical Association*. 2010;304(20):2279-2287. doi:10.1001/jama.2010.1692.

281. Carman JM, Shortell SM, Foster RW, et al. Keys for successful implementation of total quality management in hospitals. *Health Care Manage Rev*. 2010;35(4):283-293. doi:10.1097/HMR.0b013e3181f5fc4a.

282. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Quality & Safety*. 2012;21(10):876-884. doi:10.1136/bmjqs-2011-000760.