Use of Six Sigma and Toyota Production System in Health Care Quality Improvement:
A Systematic Review of the Literature

Ian B. Buchanan, M.D.

A Master's Paper submitted to the faculty of
the University of North Carolina at Chapel Hill
In partial fulfillment of the requirements for
the degree of Master of Public Health in
the Public Health Leadership Program.

Chapel Hill

2006

First Reader _____
Russell P. Harris, M.D., M.P.H.
Director, Health Care and Prevention Program
UNC School of Public Health

Second Reader _____
Lawrence K Mandelkehr, MBA, CPHQ
Director of Performance Improvement
UNC Health Care

Date _____

Table of Contents

Abstract

In the year 2000, the Institute of Medicine released a groundbreaking report that indicated that as many as 98,000 Americans die each year as a result of medical errors. This report and others that followed have created interest on the part of patients, providers, and purchasers of health care in ways to improve the quality of health care services in the United States.

Although it is clear that the health care system must undergo significant transformation if we are to improve patient outcomes, the means to achieve that transformation are not clear. Several proprietary quality improvement methods have been developed in the manufacturing industry and have been successful in improving quality in that industry. Two of these proprietary methods are Six Sigma and Toyota Production System. Recently, some health care organizations have been attempting to use Six Sigma and Toyota Production System methods to improve the quality of patient care, but evidence for their effectiveness in health care is limited.

This paper provides a systematic review of the published literature related to the use of Six Sigma and Toyota Production System in health care and, upon finding an almost complete lack of evidence for their use, considers the changes needed in the design and execution of quality improvement projects in order to be able to publish compelling studies based on the projects.

Background

Introduction

The Institute of Medicine Reports

"Human beings, in all lines of work, make errors."[1] This is among the first statements in the highly influential year 2000 report *To Err is Human: Building a Safer Health System*, published by the Institute of Medicine (IOM). This report estimated that between 44,000 and 98,000 Americans die each year in hospitals as a result of medical errors.[1] The report concluded that the primary problem was not with care givers, but with the health care system itself. As such, the report called for systemic changes related to the need for nationwide error reporting systems, free flow of information, and devotion of greater government funding to patient safety-related projects and agencies.

Despite some controversy over the accuracy of the mortality figures reported in the study, the effect of the study has been to activate stakeholders to the very real problems of medical error in our modern health care system.[2, 3] Indeed, only weeks after the publication of the report, the U.S. Congress and the President began efforts to implement the report's recommendations.[4] Such rapid action by top policymakers is an indication of the perceived importance of the findings of the report.

In 2001, the Institute of Medicine published a new report entitled *Crossing the Quality Chasm: A new health system for the 21st century*. Where *To Err is Human* focused on patient safety, *Crossing the Quality Chasm* took a broad look at the problems with health care delivery in the United States. The report outlined

six major aims for all health care provision: that it should be safe, effective, patient-centered, timely, efficient, and equitable.[5] These six aims have provided the basis for much of the quality improvement work that has been done since at every level from governmental regulation to individual doctors' offices.

Taken together, these two reports from the Institute of Medicine could be seen as accomplishing two goals: 1) creation of interest in the need for change on the part of major stakeholders including health care providers, purchasers, patients, and policy makers, and 2) outline of a vision of the ultimate goal of the health care system, with a general idea of what changes would be needed to move the system toward that vision. The influence of these two reports is evident today in the expanding volume of literature related to health care quality improvement.

What the IOM reports do not attempt to do is to provide an operational method for achieving the dramatic changes in health care that are needed if we are to "cross the quality chasm." The IOM reports intentionally stay at "high altitude," avoiding the details of how to effect system change. Thus, it is up to individual health care organizations to attempt to find ways to move toward the vision laid out by the IOM reports. Two radical, organization-level quality improvement systems that some health care organizations are beginning to embrace are Six Sigma methodology and Toyota Production System methodology. In order to understand the use of these two systems and evaluate their potential to improve the quality of health care, it is necessary to review the history of the field of quality improvement overall and quality improvement in health care in particular.

Framework for Quality Improvement Efforts. An elegant and useful conceptual model for understanding medical error was formulated as part of the IOM's National Roundtable on Health Care Quality and is presented by Mark Chassin and colleagues.[6] In his report, Chassin broadly groups medical errors into three categories: underuse, overuse, and misuse. Underuse refers to low rates of utilization of proven therapeutic strategies such as providing beta-blockers for patients without contraindications after a myocardial infarction. Overuse refers to the excessive use of a therapy or diagnostic test that does not have benefit for the patient. Finally, misuse refers to choosing the wrong therapy for a given patient or condition.

Chassin also makes the important point that poor quality care costs more than good quality care. This connection is critical, since the best efforts to improve the quality of care will fail in the long run if they do not also provide a compelling business case for the institution.[7]

Although good care probably does cost less than poor care, incentives are, however, often misaligned so that the person or organization making an investment in quality often does not realize the ultimate cost savings of the improvement.[8] Efforts to make systemic improvements in the health care system must allow for incentives to be better aligned so that investments and returns are in roughly the same direction.

By categorizing error and noting the importance of efficiency and cost-effectiveness, Chassin effectively delineates the parameters under which a quality improvement program must operate if it is to be successful in the long term and

effect substantial change in the system. The challenge for hospitals and others in health care is to find a method of improvement that conforms to these parameters.

In order to better understand the context of quality improvement in health care, it is useful to examine the history of quality improvement in health care and in other industries.

History of Health Care Quality Improvement

The Flexner Report. Current efforts at system-level change in health care are best viewed through the lens of previous change efforts. An early and dramatically successful effort at quality improvement in the U.S. health care system directly resulted from the 1910 publication of the report "Medical education in the United States and Canada." This report, completed by former teacher and school principal Abraham Flexner, was a scathing account of the then-current state of medical education in the U.S.[9] In the report, Flexner noted that the majority of medical schools had virtually no admissions standards and that many would graduate any student who could pay the tuition. The result, according to Flexner, was that physicians in the U.S. were, as a group, under-trained and unfit to practice medicine, even by the standards of the time.

Much like the IOM reports some 90 years later, the Flexner Report had the right combination of properties to attract attention and catalyze change. The report was simple to understand, had stunning conclusions, and gave clear, if broad and non-specific, recommendations for change. The report is more remarkable in light of the magnitude of system-level change that resulted from its publication. The Flexner report led to a dramatic restructuring of medical

education in the U.S., and was in large measure responsible for the structure and function of modern medical education and practice.[10]

Evidence-based medicine. Medical practice has evolved over millennia largely on the basis of observations and pattern recognition passed from one physician to another. As medical science evolved and became more complex and powerful, the demand increased for objective, scientifically-based knowledge to guide practice. This demand has resulted in the current emphasis on evidence-based medicine. Physician and epidemiologist David Sackett defined evidence-based medicine as, "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."[11] This definition has since been updated to include other factors such as patient values but the basic idea is the same: medical practice should be based on information acquired in a systematic method that reduces bias and increases the likelihood of finding the "true" answer to a clinical question.[5]

Evidence-based medicine is now a fundamental and widely accepted pillar of medical practice. This assertion is validated by the huge number (over 131,000 in the last 50 years) of clinical trials published in recent decades.[5] Acceptance of evidence-based medical care is now wide-spread among physicians and is a fundamental part of undergraduate and graduate medical education.

Widespread understanding and acceptance of the need for a scientifically-sound knowledge base for clinical practice has not translated well into the realm of health care quality improvement. The majority of physicians have no formal training in quality improvement methodology, and thus are left to "reinvent the

wheel" if they want to improve patient care or, worse yet, not engage in quality improvement at all. Furthermore, use of established quality improvement practices is hampered by factors similar to those that reduce the use of evidence-based medical practices: difficulty of integration with practice, clinical priority-setting, and suspicion of the accuracy of findings.[12]

Defining and Measuring Quality. Evidence-based medicine is built on the principal that one can measure the effect our actions have on patients and then compare the relative effects of different actions. It is relatively easy to measure mortality, cholesterol level, or blood pressure. It is far more difficult to measure, or even define, quality. Despite the difficulty, defining and measuring quality is exactly what is required if we are to approach quality improvement in a systematic, methodologically rigorous manner.

One of the most influential figures in defining and measuring quality, particularly as it relates to health care, is Avedis Donabedian. Dr. Donabedian has spent a career thinking and writing about health care quality. His definition of quality focuses on processes of care and the extent to which those processes contribute to desired health outcomes. Donabedian further outlined a method for quality assessment based on three broad, interrelated categories: process, structure, and outcome.[13] The process of care includes all aspects of how care is delivered and the interface of the patient, the provider, and the health care system. Structure refers primarily to the physical attributes of a health delivery system (e.g., adequate building space, proper laboratory facilities). Outcome refers to actual measured patient outcomes such as mortality and morbidity. These basic

definitions and measurement parameters for health care are the basis for many of our current quality assessment activities, such as those of the Joint Commission on Accreditation of Healthcare Organizations and the National Committee for Quality Assurance.[14, 15]

Donabedian's work has provided us with an operational definition of quality and a template for measurement of quality. This template can form the basis for a system of quality improvement with standard metrics that allow comparison of interventions, and thus facilitate the use of standard techniques for quality improvement.

With renewed interest in system-level health care improvement in the wake of the IOM reports, the time has come to look for standardized, proven methods for bringing about quality improvement in a health care organization. In the same way that physicians are not expected to come up with their own treatment methodologies for hypertension, they and others in the health care field should not be expected to come up with their own means of improving quality. The use of a proven, formal methodology for quality improvement could be a great boon to health care practitioners and organizations. Six Sigma methodology and Toyota Production System methodology are two such methods. Their use in industry and evidence for their value in health care are the focus of the remainder of this paper.

Managing for Quality

Six Sigma and Toyota Production System are both methods that rely on the idea that quality is something that can be managed and "built-in" to a system

or organization. In order to understand how these systems work, one must understand their theoretical underpinnings, which rest on the work of two pioneers of quality improvement: Walter Shewhart and W. Edwards Deming.

Shewhart and Quality Control. Walter Shewhart, a physicist, statistician, and employee of Bell Laboratories in the mid-20th century, was among the first to propose a system of continuous measurement for the purpose of improving the quality of manufactured products and lowering their defect rate. Prior to Shewhart, the manufacturing industry typically relied on inspection-based quality control. In an inspection-based quality control system, finished products are inspected for defects that, if found, result in the product being reworked or discarded.

Shewhart recognized this process as inherently inefficient, and developed a monitoring system where the numbers and types of defects were recorded over time and the results were plotted on a "run chart" in which the x-axis was time. This allowed managers to see when more or fewer defects were occurring. Shewhart further refined his method by applying statistically based "control limits" to the charts, which allowed managers to determine if variation in defects was due to "common cause" (minor, random variation), or "special cause" (major, identifiable events/factors that could be corrected).[16] This process allowed, for the first time, managers to correct sources of defects and thus prevent future defects related to the same cause. It also removed much of the need for costly, inefficient inspections at the end of the manufacturing process.

Deming and a System of Quality Improvement. While Shewhart put forth the basic mechanism for tracking quality and identifying causes of variation, more was needed to come up with an actual approach to improving a process once problems were identified. The individual with the greatest influence in this was W. Edwards Deming. Deming recognized the interplay between the production system and the workers, and created a thought framework that dictated that quality improvement efforts had to be accessible to the "front-line" worker and that, furthermore, these workers must "buy-in" to the process of quality improvement.[17]

Deming's work was most widely accepted and integrated into industry in post-World War II Japan. In Japan, industry leaders embraced the notion that quality and quality management should be handled, as much as possible, at the level of the line employee. Japanese corporations moved away from inspection-based quality control, and moved toward a model in which each employee took responsibility for the quality of the product, and was empowered to halt the manufacturing process at any point if a defect was recognized. This allowed defects to be caught earlier in the process, at a point when fixing the mistake was relatively simple and inexpensive.

Eventually, Shewhart's and Deming's work was synthesized into The Model for Improvement that forms the basis of much of today's industrial quality improvement processes.[18]

The Model for Improvement. The Model for Improvement is based on the goal of answering three questions: 1) what are we trying the change? 2)

how will we know the change is an improvement?, and 3) what change can we make that will result in improvement? Answers to these questions come from a protocol Deming developed called the Plan, Do, Study, Act (PDSA) cycle. This cycle was a modification of the process described by Shewhart as the Plan, Do, See process.[19]

The PDSA cycle is analogous to the scientific method in that one starts with a hypothesis (part of the Plan step), designs and implements an intervention (Do), observes predefined process and outcome measures for signs of change as a result of the intervention (Study) and finally decides to keep, abandon, or modify the intervention (Act). PDSA cycles are designed to be simple, generally addressing only a single aspect of a problem. This simplicity is intentional, as it allows cycles to be completed rapidly and for new cycles to begin that are based on the results of previous cycles.

Summary. The work of Shewhart and Deming was adopted rapidly and with almost religious devotion by the Japanese in the wake of World War II. Japanese industry is built on the principles of these quality pioneers. It was the work of Shewhart and Deming that inspired the creation of the Toyota Production System. Widespread use and acceptance of these methods in the U.S., however, would not occur until the 1980's. It was around this time that Japanese manufacturers began to be seen not simply as low-cost, low-quality sources of toys and other cheap items, but as powerhouses of efficiency and quality that were capable of challenging and besting U.S. manufacturers for market share.[20] Under this market pressure, U.S. manufacturers rediscovered the principles of Shewhart

and Deming and corporations began to devote substantial time and energy not only to implementing the lessons but creating their own spin on the principles. These investments would lead to proprietary systems of quality improvement such as Six Sigma methodology.

Six Sigma in Industry

History and Development. The term "six sigma" relates to the statistical concept of six standard deviations from the mean (with the Greek letter sigma (σ) representing a standard deviation). Mathematically speaking, out of one million "events" arrayed in a normal distribution, 3.4 "events" will lie outside of the six sigma mark.[21] In other words, achieving "six sigma" performance in a process implies an occurrence of 3.4 or fewer defects per million opportunities (DPMO) for a defect. Many major manufacturers today recognize six sigma quality as a goal that is both desirable and achievable in their manufacturing processes.

In an effort to improve its manufacturing processes and its competitive edge, Motorola Corporation developed a proprietary quality improvement process called Six Sigma in 1986.[22] The initial goal of the project was to develop a systematic quality management technique that would allow Motorola to achieve six sigma quality in its manufacturing.

The quality improvement process developed at Motorola is, as noted earlier, an extension of the PDSA cycle developed by Shewhart and Deming. Under Six Sigma methodology, the PDSA cycle is modified to DMAIC: Define the objectives and parameters of the project, determine what Measures will be

used to track progress, Analyze how the process functions and where improvement can be attempted, Improve the process, and Control the results of the intervention to ensure the business case is met and that the process is improved[23]. Along with the use of the DMAIC process, another core tenet of Six Sigma is that its use must be supported by the top levels of management. Six Sigma is not designed as a tool to modify day-to-day processes. Rather, it is designed as a tool to facilitate large scale, strategic system change. Such change in organizations is only possible with support from the highest management levels.[24]

The final major element of the Motorola Six Sigma process is the training of employees to use the Six Sigma process. Training culminates with certification of an employee at one of four levels, analogous to martial arts achievement levels. Employees receive White Belt certification after some training and they can help facilitate projects. Green Belts can assist with managing large projects and can be in charge of small improvement teams. Black Belts are individuals who have extensive training and experience with Six Sigma and are the people responsible for running Six Sigma projects. Finally, Master Black Belts are individuals who are qualified to oversee all of the projects in an organization and provide support to the Black Belts.[25] Training an organization's employees to engage in Six Sigma projects is a way to bring quality to every "front-line" worker, and reinforces the Six Sigma tenet that quality must be an organizational culture shared by all workers, regardless of occupational level.

Application of Six Sigma. The underlying assumption behind achieving six sigma quality is that such efficiency will save a corporation money and increase revenue, even though the financial investment in the Six Sigma process is substantial. In the corporate setting, this assumption has been shown to be correct at Motorola and at other corporations who have implemented the Motorola Six Sigma methodology. Motorola reports that since implementation of Six Sigma in 1986, they have saved over $17 billion in their manufacturing processes.[26] Furthermore, early adopters of Six Sigma such as General Electric report increasing their operating income by $300 million in the first year of Six Sigma's use.[27] This magnitude of return on investment is the result of a lower defect rate, more satisfied customers, and competitive edge in the marketplace. For manufacturing companies, this is all the proof that is required that Six Sigma is a valid and reliable method for quality improvement. In health care, where quality measures of, "patient function, patient satisfaction, clinical outcomes, employee satisfaction, and cost," all tend to come before market share and profit margins, the applicability of Six Sigma is less clear.[28]

Toyota Production System in Industry

History and Development. The Toyota Production System (TPS, sometimes used interchangeably with the term "lean production") was developed, not surprisingly, at Toyota Motor Company. Toyota is a world leader in the manufacture of automobiles, with a reputation based largely on the superior quality and reliability of their products.[29] Where Six Sigma is a highly formalized process designed to provide the framework and tools for systemic organizational

improvement, TPS is more a philosophy or set of principles that are designed to be integrated into an organization's work in such a way that each work process directly provides value to the customer. TPS is built on the idea that there is one and only one correct way to do any given job, and that all work tasks should be clearly defined as to how and when they are performed, and by whom. Change and improvement in processes come about because the only way to change the way a process is performed is to suggest and test an alternate method.[30]

Implicit in the TPS method is that workers must have not just the right, but the obligation, to 1) stop a process as soon as a defect or possible defect is noted, and 2) work to find new and possibly better ways to perform given tasks. Complete acceptance of this set of principles was in large measure responsible for Toyota's dramatic success in Japan and the United States, as well as its ongoing success despite changing market pressures.[20]

Although Six Sigma certainly requires involvement and support from the top levels of an organization in order to be effective, TPS requires an even greater integration into an organization's culture. Few companies have been able to replicate the success that Toyota has derived from TPS. This may be due to an incomplete understanding of the depth to which the system must permeate the culture of the organization.[31]

Health care providers are often rankled by the notion of using an auto company's production principles to improve health care. Nevertheless, although patients are certainly not equivalent to automobiles, the driving force in each industry is the same: to provide the highest quality product for the lowest cost.

TPS has as its central tenet that all production systems should operate with the customer in mind, and that activities that do not add value for the customer should be eliminated.[32] When parsed in this context, the potential for TPS to be used in health care is more apparent.

Application of TPS. TPS has been implemented to a greater or lesser degree by organizations that span numerous industries. Some examples of TPS success beyond Toyota include the Canada Post Corporation (CPC, equivalent to the U.S. Postal Service). The CPC implemented TPS at every level of operations, and over the first few years after implementation achieved a $300 million return on investment, along with improved service including a 28% reduction in transit time for mail.[31] Another example of successful use of TPS is found in Genie Industries, manufacturer of lift devices. Genie has implemented TPS and seen annual total costs decrease by 5% each year since implementation.[31] Similar to that of Six Sigma, the success of TPS in industry can generally be quantified in terms of cost and profitability. TPS utility in health care is more difficult to measure.

Use of Six Sigma and TPS in Health Care

As stated in the Introduction, there is little doubt that health care could benefit from major system changes to reduce the rate of errors and other "defects." While manufacturing corporations have reduced their defect rate to below the six sigma threshold, many health care organizations and processes routinely produce one sigma results, equivalent to 500,000 DPMO. By comparison, airline baggage handling results in 4,000 DPMO (4.1 sigma) and

airline fatalities occur at a rate of 0.4 DPMO (>6 sigma).[33] Most people would like to think that the chance of a medical error is less likely than the chance of lost baggage, but clearly this is not the case. With this in mind, it is useful to examine the types of changes organizations are making that employ Six Sigma and TPS methodology.

Six Sigma Applications in Health Care. When speaking of the use of six sigma in health care, it is useful to be precise as to whether one is speaking of efforts to achieve six sigma quality (i.e., 3.4 DPMO), or whether one is speaking of utilizing the Motorola-developed Six Sigma methodology. For the purposes of this review, six sigma *without capitalization* will denote aiming for 3.4 DPMO, while Six Sigma *with capitalization* will denote the use of the Motorola methodology or a derivative proprietary methodology.

Six Sigma has been used by numerous health care organizations, for a variety of purposes. Some of the health care service areas in which Six Sigma-based improvement has been tried are discussed below.

Laboratory and Pathology Services. Medical laboratories have a long commitment to quality improvement, with formal efforts at standardization and improvement dating at least as far back as the 1950s.[34, 35] Laboratories have willingly embraced mandates for quality such as the Clinical Laboratory Improvement Amendments of 1988, as well as voluntary, professional society sponsored quality efforts like those of the National Committee for Clinical Laboratory Standards and the International Organization for Standardization.[36] Perhaps the laboratory is more amenable to quality improvement since its

operations are seen as more standard and less subject to variation than patient encounters or other clinical activities. A recent study of auxiliary services at a Naval Medical Center found that many laboratory services already operate at between four and five sigma levels of error.[37] Such performance would be considered exemplary in any health care field, but efforts are underway to improve this level of performance with Six Sigma methodology.[33]

Radiology. Several published articles discuss the use of Six Sigma methodology in radiology. The use of Six Sigma in radiology is perhaps not surprising considering that GE Medical Systems is not only a major supplier of imaging equipment, but also a supplier of consultants to run Six Sigma projects for hospitals.[38]

The radiology department at the M.D. Anderson Cancer Center in Houston, Texas, used Six Sigma to reengineer their film library to provide substantial improvements in film tracking and film availability to clinicians.[39, 40] In similar fashion, the radiology department at Stanford University has used Six Sigma to improve availability of appointments for CT scans, and in the process has increased capacity and projected revenue from the scanners.[41] Each of these organizations has reported glowing success with the use of Six Sigma methods to improve their internal and external performance.

Pharmacy. A number of health systems have made efforts to implement Six Sigma projects that decrease medication errors and improve pharmacy services. Cardiovascular Surgical Associates, a large group practice in Kentucky, implemented a Six Sigma project focused on improving pharmacy services.

Through Six Sigma, the group implemented an e-prescribing solution that decreased patient and pharmacy callbacks related to prescriptions and that improved efficiency of operations.[42] Patient care was improved by virtue of less confusion on the part of patients and pharmacists related to prescriptions.

Other organizations report having used Six Sigma to improve the process of IV medication administration and reduce errors in administration of medications.[43] The use of Six Sigma in pharmacy services is still early in its development, but roadmaps do exist that could help other organizations implement Six Sigma successfully.[44]

TPS Applications in Health Care. There is less published information available about the use of TPS in health care than about the use of Six Sigma. This may be related to the relatively greater institutional commitment required for TPS compared to Six Sigma, or it may be related to TPS lacking a major health care corporate advocate, as Six Sigma has in GE Medical Systems. However, a few institutions have published their experiences with TPS.

Toronto's Hospital for Sick Children implemented TPS in the management of their radiology department and has since reported a cost savings of $140,000 over 4 years in their interventional radiology services.[45] The University of Pittsburgh Medical Center has embraced TPS throughout the organization, and has reported promising results in administrative areas like nurse retention and in clinical areas such as medication error rates and hospital acquired infection rates.[30, 46]

Summary of Health Care Use of Six Sigma and TPS. In addition to the projects discussed above, a multitude of other Six Sigma and TPS projects have been undertaken in health care in disparate areas such as pressure ulcer care,[47] billing activities, [48] nurse retention,[49] and emergency department operations.[50] Even more numerous than the publications citing specific projects and their results are articles that shout the virtues of Six Sigma and TPS for health care, without providing evidence of results derived from the type of rigorous evaluation methodology that is required of current biomedical research.[23, 25, 51-55]

Given the number of health care organizations that are beginning to embrace Six Sigma and TPS as primary means of system redesign, and the number of articles available that discuss the use of these methods in health care, I felt a systematic review of published literature on these quality improvement methods was needed. The need for a systematic review arises partly from recognition that the majority of available articles are of low methodological quality and partly from the sense that, because Six Sigma and TPS require huge commitments of time, money, and resources in organizations where they are used, good evidence needs to be available before institutions make full commitments to these systems.

Methods

Defining the Key Question.

The U.S. health care system is in need of radical efforts at systemic change, as evidenced by the IOM report and by other reports that indicate that patients routinely receive poor or inappropriate care.[56, 57] Quality in health care

has been sufficiently defined by thought leaders like Avedis Donabedian and the components of quality generally agreed upon by the major stakeholders. The tools have been developed to track errors and defects in medical care and analyze the results to find reasons for the defects. Finally, these techniques have been adapted by major manufacturing companies and developed into proprietary protocols and processes for improvement that have demonstrated clear success within the industries in which they were developed.

The fundamental question is whether these techniques are appropriate for use in the health care industry and whether they can provide benefits in the form of decreased medical errors and improved patient care, while meeting the requirement of a favorable business case for their use. This key question is framed in the context of the relevant population, intervention, comparison group, and outcome.

Population. The population of interest for study is, indeed, the entire health care system. Interventions are implemented on a smaller scale, such as an individual hospital department or nursing floor, but the results should be meaningful on a larger scale related to entire hospitals or the health care system as a whole.

Intervention. The intervention of interest is implementation of Six Sigma or Toyota Production System methodology in an effort to improve quality. For the purposes of this review, this implementation could relate to business processes or patient care processes, since both potentially contribute to improved patient care and/or satisfaction.

Comparison. The results of implementation of Six Sigma or TPS techniques should be compared to "business as usual," whether this is usual processes of patient care or usual business practices.

Outcome. Outcomes are defined primarily by some metric of improved quality, which would vary depending on the purpose of the intervention, or by cost savings, either to the institution or to the patient. Since cost is so intimately related to quality, it is valid to consider cost savings a quality improvement, as long as the savings are the result of improved processes rather than provision of substandard patient care.

Inclusion/Exclusion Criteria. Together, the criteria outlined above form the inclusion criteria for articles to be included in the systematic review. An article must have relevance to the health care system, have either Six Sigma or TPS and its intervention, have some reference to a comparison group, and have explicitly defined outcomes.

Exclusion criteria are: failure to meet all of the above inclusion criteria, as well as failing to formally denote study design, duration of the study, and data analysis methods.

An argument could be made to include in the systematic review studies in which the primary intervention is described as "lean production" methods. This methodology is closely related to TPS and in some cases the terms are used interchangeably. However, I have chosen to exclude these studies from this review because of the heterogeneity of what different authors define as lean

production. This heterogeneity makes it impractical to consider lean production as a single intervention.

Systematic Review Method

Literature Search. I searched MEDLINE (directly and through the PubMed interface), The Cochrane Collection, and the University of North Carolina Library Catalog to identify articles, books, and manuscripts relevant to health care quality improvement and specifically relevant to the use of Six Sigma and Toyota Production System methodology in the health care setting. Searches were conducted using the following search terms, individually and in appropriate combinations: "Six Sigma," "Toyota," "quality improvement," "health care," and "healthcare." I limited searches to English language publications. Searches were not limited according to date of publication.

I used National Library of Medicine publication meta-data to identify any randomized controlled trials, meta-analyses, or review articles. Relevant abstracts were imported into the EndNote 9 citation manager.

My searches revealed 1732 possible references. Out of this number, I excluded, based on the article title, 1643 references for lack of relevance to the review subject or for lack of an abstract for review. Of the remaining 131 titles, I was unable to obtain the abstract for another 16 titles. I excluded 42 articles on the grounds that they turned out to be foreign language, were off topic, or were editorials. I retrieved full text for the remaining 73 articles. These articles consisted of a mix of background articles, "white papers," case reports, and prospective studies. One case report and four prospective studies met the initial

criteria for inclusion in the review, which included a clear intervention, use of explicitly defined metrics, and formal data analysis. A full QUORUM tree is available as Appendix I.

Study Selection. As noted above, very few published articles exist that describe the evaluation of Six Sigma or TPS in a formal, scientifically rigorous manner. As a result of the severe limitation of the current literature, I have included all five articles that met initial inclusion criteria in the systematic review.

Data abstraction. I used a data abstraction form to increase consistency of data abstraction for the five articles included in the review. The form recorded the study design, duration, location (both institution and unit, if applicable), intervention (either Six Sigma or TPS), outcome measures, main results, and limitations, including sources of confounding and bias. I assigned each study a subjective rating of internal validity based on the strength of its methodology. This rating is based on the rating system of U.S. Preventive Services Taskforce (i.e., good, fair, poor). In addition, I assessed the external validity of each study.

Results

Evidence for Use of Six Sigma in Health Care

Four of the five studies that met criteria for inclusion in the systematic review addressed the use of Six Sigma methodology in a health care setting. Despite meeting the basic inclusion criteria for the literature review, none of the four studies can be considered high quality. The common failing among all of the studies is that they utilize a time series design, in which there is a single group that is the target of the intervention and outcomes for that group are compared to

historical outcomes from the same group. Although this design is the most pragmatic to use when evaluating the success of a quality improvement intervention, it does not provide adequate safeguards against confounding and other sources of bias to make it a reliable scientific design. The objectives of the four studies ran the gamut from hospital operational improvements to direct patient care projects. I will briefly discuss the studies individually in order from most- to least-flawed.

Chan and colleagues report on the use of Six Sigma methodology to analyze and decrease errors in prescription medication dispensing in an outpatient clinic pharmacy.[58] The main outcomes of the study were the number of medication errors detected per million prescriptions. The authors reported a rate of 338.8 DPMO before the intervention and a rate of 230 DPMO afterward. There is no statistical comparison of the results. Further, there is high potential for confounding, since the participants were aware of the study and may have performed differently as a result. In addition, the authors do not address the potential for errors to go unnoticed, and thus uncounted. Because of its poor methodology, it is impossible to determine the degree of generalizability of the results.

Kang and colleagues report on the use of Six Sigma to increase the efficiency of maintenance of a Picture Archiving Communication System (PACS) digital radiology system in a hospital in Korea.[59] The main outcome measures are the "actuarial halts per time" (a measure of the impact a system component failure has on the system) and the resource utilization level required to keep the PACS

system functioning. Strengths of the study are that the authors rigorously define their outcome measures and provide a clear discussion of what changes were made and what effects they had on the system. The authors found that most aspects of the system were already performing at near-six sigma levels. They used Six Sigma techniques to redesign their maintenance so as to decrease the resources required to keep the system running. They report that the redesign allowed for maintenance to be performed at 79% of pre-intervention levels. Unfortunately, they do not discuss what effect the changes had on the actuarial halts per time, so readers are left to assume that it did not increase as a result of the changes. In addition, since the people doing the recording of maintenance logs were the same as the people doing the study, potential exists for measurement bias in the outcome measures.

Adams and colleagues report on the use of Six Sigma to decrease operating room turnover times in their hospital.[60] The main outcome measure of the study was overall turnaround time, including its component parts of surgeon-out to patient-out, patient-out to patient-in, and patient-in to surgeon-in. The authors found that use of Six Sigma techniques resulted in a decrease in the Z score (a measure of the proportion of cases violating the upper control limit for turnaround time) from 2.13 to 1.53. Possible issues with the study include the use of nurses to record turnover times, since these nurses were aware of the project and could be biased. This concern is somewhat allayed by the use of a validation sample of 20 cases in which two nurses recorded time and the times were then compared for reliability. The authors found a 12 minute discrepancy between

observers in over 17 hours of observations. Aside from the inherent limitations of the study design, the internal validity of the study seemed reasonable. Further, the study results are probably reasonably generalizable, since many hospitals face a similar problem with their operating rooms, and have similar processes in place.

The best of the four studies relating to Six Sigma is from Frankel and colleagues who report on the use of Six Sigma to decrease the rates of catheter-related bloodstream infections (CR-BSI) in their surgical ICU.[61] The main outcome measure in the study was number of CR-BSIs per 1000 catheter days. The authors found that the CR-BSI rate decreased from 11/1000 catheter days to 1.7/1000 catheter days ($p<0.0001$) over the course of the study. The primary limitation of the study is, again, its time-series design. In addition, since the staff in the ICU were aware of the study, their performance may have changed as a result of the study rather than the intervention. Also, there was no effort to determine whether any change in the patient mix in the ICU was different during the study period compared to before. Despite these problems, the results of the study are sufficiently dramatic to indicate that the intervention likely did have a positive effect on the outcomes.

Evidence for Use of TPS in Health Care

As scant as published evidence for the effectiveness of Six Sigma techniques in health care is, the evidence for use of TPS is even less. I found only one article on TPS that met even the basic inclusion standards of this review. Raab and colleagues report on the use of TPS to decrease the rate of inadequate specimen collection for Papanicolaou (Pap) testing.[62] The main outcome

measures in this study were the frequency of absence of transformation zone cells (as a definition of an inadequate sample), and the percentage of tests diagnosed as ASC-US (also a measure of inadequate sample collection). The study found a decrease in frequency of samples without transforming zone cells from 9.86% to 4.74% (p=0.001) during the study. In addition, the authors found a decrease in ASC-US frequency from 7.8% to 3.9% (p=0.008). The study has many problems. The most glaring is that the intervention was used with only one gynecologist, and he was self selected and described in the study as "expressing enthusiasm about improving his Papanicolaou test sampling."[62] In addition, the study authors seem not to have a clear sense of the study design, referring at various points in the article to "cases and controls" and "preintervention and intervention," all to refer to the same observations of the participating gynecologist. The study is so flawed that it would be difficult to make any generalizations about the use of TPS to improve Pap sampling, much less any other health care activity.

Summary of Literature Review Results

It is clear based on the quality of the studies discussed above that the literature is woefully lacking with respect to the use of Six Sigma and TPS methods in health care. A summary evidence table of articles included in the review is available as Appendix II. The primary failing of all of the studies currently available on the use of Six Sigma and TPS is that the projects seem to have been planned only as internal improvement projects, without consideration for the extra degree of methodological rigor necessary to make an intervention and its outcomes suitable for publication and, more importantly, compelling

enough to convince people not already familiar with the techniques that they work and are worth pursuing.

Furthermore, while a formal funnel plot of the available literature is not practical given the small number of studies available for review, it is clear that the bias in this literature, as with much of the biomedical literature, is toward publication of "positive" studies. I did not encounter a single published work that described a failed effort at using Six Sigma or TPS in health care. I did encounter articles that made vague references to failed projects, so there is little doubt that failure when pursuing quality improvement through Six Sigma or TPS is a real concern. Some of the obstacles to success include difficulty with acquiring the volume and type of data needed for projects and difficulty getting adequate involvement from clinical staff.[63]

Discussion

Limitations of Currently Available Literature.

Extensive search of the published literature revealed nearly 100 articles about the use of Six Sigma and TPS in health care. Out of these articles, only five met the basic methodological criteria that most clinicians would expect in order to consider an article a scientific study with results that could be trusted. Of these five, none had sufficiently rigorous methodology to be considered a highly valid study. The reasons for such a complete lack of compelling research in this field are not clear, but the deficiency may have to do in part with differences between clinicians and quality improvement professionals and their respective priorities.

Although Six Sigma and TPS are some of the most recent examples of health care borrowing improvement techniques from industry, in fact the entire field of health care quality improvement has grown out of improvement efforts in other industries. With this in mind, it is perhaps not surprising that a large percentage of full time quality improvement professionals working in health care have educational and experiential backgrounds in business and engineering. At the same time, individuals with medical training receive little exposure to formal quality improvement methods. Furthermore, under our current health care payment structure, there is no way to compensate clinicians for quality improvement activities.[64]

It would be both incorrect and unfair to imply that quality improvement professionals do not know how to perform valid studies with measurable results. On the contrary, these professionals are very good at performing such studies of change. However, pragmatism must often take precedence over rigor when designing heath care improvement projects. With this in mind, it is worthwhile to consider a structure for bridging the gap between quality improvement professionals and clinicians.

Proposed Strategy for Improving the Body of Improvement Literature.

A recurring theme in the quality improvement literature in general, and in relation to Six Sigma and TPS in particular, is that quality improvement must be integral to the thinking of every member of an organization. In the case of health

care, some of the most influential members of any organization are the clinicians who actually provide care. If we are to make systemic changes in health care, clinicians must be convinced not only of the need for change, but also of the effectiveness of the tools for change.[20] With this in mind, I will lay out a proposed method of design and implementation for quality improvement projects that would have sufficient methodological rigor to conform to the expectations of clinicians.

Project Selection. In order to be able to carry out and publish a quality improvement project that will garner the buy-in of clinicians, careful consideration should be given to which projects have relevance to clinicians. Even the most methodologically rigorous quality improvement project might have little influence on a clinician's view of a given improvement methodology if the topic does not relate to clinical care. In many institutions, quality improvement departments are responsible for carrying out projects related to non-clinical organizational functions. These projects are probably not suitable for publication if the goal is to influence clinician thinking. One way to identify projects that would provide material for a compelling publication is to involve clinicians from the earliest stages of a project. It is well established that clinician involvement in quality improvement projects is critical to the long-term success of the projects.[65, 66] In this same vein, clinician involvement could be thought of as essential to the identification of quality improvement projects that are of great enough clinical relevance to be worth publication.

Improvement Tool Selection. Clinicians are, by nature, skeptics when it comes to accepting change based on new data. In order to convince clinicians of the value of an intervention, data must be collected and analyzed so as to make it "bulletproof."[67] To maximize acceptance of improvement data, it is important to design improvement projects that conform as much as possible to the scientific method with which clinicians are familiar and comfortable. One tool with strong parallels to the scientific method is the PDSA cycle. Use of the PDSA cycle as the basis for design of quality improvement projects has several advantages. First, the PDSA cycle is already used and understood by virtually all quality improvement professionals. Second, the PDSA cycle is easily explained to individuals without experience in its use. Third, direct connections between PDSA and the scientific method can be explained to clinicians and others.[68] Fundamentally, the PDSA cycle defines a problem, lays out a discrete intervention, and delineates measures that will indicate whether the intervention resulted in a change in a defined process or outcome. This is directly analogous to the scientific method, and easily understood by clinicians. Although the PDSA cycle is a useful, scientifically based underpinning for successful project design, its use is not sufficient to make a project scientifically rigorous. Thought must be given to collecting data in a way so that valid conclusions can be drawn about the effect of the PDSA cycles or of other interventions.

Study Design. The single biggest weakness with all five published studies that met criteria for inclusion in this review is the validity of the comparison of a process before and after the intervention. These studies all used

a time series design (also called a before-after design). This design has been used in quality improvement in industry for decades, and is probably sufficient in that environment, where outcomes are relatively obvious and easy to measure and, importantly, where confounders are less of a consideration due to the controllability of the manufacturing process. The design is less well suited for application in health care, where outcomes are often more difficult to both define and adequately measure. Unfortunately, quality improvement professionals often feel that other approaches to comparison group creation are too logistically difficult in the health care setting.[69] Nevertheless, if we are to improve the literature on quality improvement, it is important to move beyond the before-after design and attempt more rigorous designs that resemble the best examples of biomedical study design.

Use of Randomized Controlled Trials. In some instances, it may actually be possible to conduct a randomized-controlled trial of a quality improvement intervention. In these cases, proper selection of "study subjects" is crucial to ensure validity of results. Because most quality improvement focuses on system-level changes, an individual study subject might in fact be an entire nursing floor or even a hospital department. For this reason, enrolling enough "subjects" to achieve adequate statistical power becomes a major concern. One innovative way to increase the pool of potential "subjects" is to involve several organizations in the improvement effort.

One of the most remarkable examples of multiple hospitals working together on quality improvement projects is the Veteran's Administration (VA)

hospital system. The VA has achieved remarkable quality improvement over the last decade and now in many respects provides the highest-quality health care in the United States.[70] This dramatic turnaround has been due, in part, to the ability of multiple VA hospitals to work together on quality improvement projects, greatly enhancing the power of a project to evaluate the success of improvement interventions.[71] Although the VA has a clear advantage in creating inter-hospital collaborations by virtue of being under a single, central command, other hospitals have begun to form collaboratives to enhance the power of their improvement efforts. A number of hospitals in the Pittsburgh area have banded together in just such an effort, and are beginning to make progress on achieving patient safety goals.[72] Also, leading quality improvement organizations, such as the Institute for Healthcare Improvement, have developed mechanisms for hospitals to work together on improvement efforts, with the most notable example being the Breakthrough Series Collaborative.[73] Collaborative efforts among institutions could make the performance of randomized quality improvement projects possible, while also speeding the spread of improvement knowledge.

Using Non-Randomized Study Methods. When randomized controlled trials are not possible, other options exist that are superior to the standard time series study design. The time series design of many improvement projects is subject to a number of biases including maturation (change in the process unrelated to the intervention), history (events occurring during the study that influence outcome but do not relate to the intervention), and testing (change in the system related to the process of being studied, rather than to the intervention) to

name just a few.[74] One modification of the time series design that can compensate for many of the biases of the time series is to use a so-called ABAB or equivalent time series design. In this study design, measurement of the outcome of interest begins at baseline and continues throughout the study period. The intervention of interest is introduced for a period of time, then removed, then reintroduced, then again removed. The idea is that if the intervention is actually having an effect, a change in the measurement should be noted after each introduction, with reversion toward the baseline after the intervention is removed each time.[69] Another enhancement to the time series design that can improve its reliability as an assessment tool is the collection of adequate data on the functioning of a process before the intervention of interest. Often, these data will have already been collected, and will form the basic impetus for the project in the first place.

In the past, improvement projects have been limited by the inability to collect enough data on the functioning of a process to establish a reliable baseline. Thus, after performing an intervention, a comparison would have to be made to only a few pre-intervention data points. The development and use of large clinical databases represents a potential boon to quality improvement projects, in that these databases can provide "instant access" to a wealth of data to inform both project selection and design.[75] At the same time, improvement professionals must be cautious in the use of large databases, since their data can be compromised or can provide a misleading picture of actual clinical processes.[76] As these databases become more prevalent in health care organizations,

measuring the "before" and "after" performance of a process will become easier and thus improvement studies will be able to utilize more extensive data comparisons.

Use of Qualitative Methods. In an interesting counterpoint to the usual demand for greater and greater use of quantitative methods in health care, at least one author has advanced the notion that in certain instances, the use of well designed qualitative methods may actually be superior. Their rationale is that in highly complex and heterogeneous systems like health care, qualitative exploration of a few "signal cases" can provide greater quality improvement insights than quantitative data collection of an entire process.[77] Despite the potential utility of this approach, it is not clear how easily clinicians would accept the results of such investigation.

Data Analysis. It is important to assess the process of interest before and after any intervention not only using classic descriptive statistics, but also using process-based statistical methods such as control charts. Control charts are constructed by plotting the performance of a system on a chart where the x axis is time. In doing so, a plot emerges that shows fluctuation around some mean value for the process. One can then calculate control limits for the process, defined as values three standard deviations above and below the mean value. The three standard deviation limit is designed to minimize the risk of both Type I and Type II error in the subsequent process analysis.[78] Once a control chart is created for a given process outcome or function, it can be used to assess whether the process is stable, or whether it is subject to large variations. Any values that fall outside of

the control limits of the chart are considered to represent "special causes" of variation that prevent the process from being stable. These special causes should be dealt with before the implementation of a quality improvement project, because they make the system unpredictable and thus make determining the effect of an intervention impossible.

Eliminating special cause variation in order to achieve a stable process before beginning an improvement project can be thought of as being analogous to selecting a study population for a traditional cohort or case-control biomedical study. Participants are generally selected on the basis of certain criteria that minimize uncontrolled variation between the study populations. In the same way, ensuring that a system is stable before a project begins minimizes variation in the process, allowing more valid conclusions to be drawn about the effect of the intervention.

Use of Statistics. Once a project of adequate methodological rigor has been completed, the final step in making the research credible to clinicians is the presentation of the results. Clinicians have grown accustomed to study results presented with complete statistical analyses. In many cases, it will not be sufficient to simply present a before and after performance number. At the same time, many quality improvement professionals are not comfortable with using statistics on the results of PDSA cycles, which, indeed, are individually not appropriate for traditional tests of significance. However, providing simple tests of statistical significance is critical if a study's results are to have credibility with clinicians.[79] Although these tests cannot be completed on individual PDSA

cycles, it is appropriate to complete them on the overall results of a project to compare the process at the end to the process at the beginning. Fortunately, clinicians are beginning to appreciate the value of control charting and other process measurement tools.[80, 81] In order to be successful, quality improvement efforts should begin to appreciate the use of p-values and other tests of significance.

Publication of Results. Finally, studies that are of high quality and clinical importance must escape from the realm of the numerous quality and management journals and be published in journals that clinicians read.[65] No matter how important a quality improvement finding is, it will not be widely recognized by busy clinicians unless it is in a journal that they routinely encounter. This final challenge is perhaps one of the greatest for quality improvement professionals, and is another area in which clinician collaboration on projects is critical. The clinician can suggest potential journals for publication and assist with appropriate manuscript preparation.

Future Considerations for Quality Improvement in Health Care.

Importance of System Change. In the last several years a major advance in health care quality improvement has been the increasing recognition that errors should, in the vast majority of cases, be thought of as system problems rather than problems with an individual person.[82] This paradigm shift in thinking, although incomplete in health care, is a critical prerequisite if we are to substantially improve health care. Once health care professionals accept this

system-based view of health care and medical error, we can make meaningful steps toward addressing a multitude of problems. All current quality improvement tools and techniques rely on having a solid understanding of how systems operate and using this understanding to change the system and reduce error. This idea of "error proofing" a process or system is a fundamental tenet of the Toyota Production System, where it is referred to by its Japanese name poka-yoke.[83] The idea of error proofing must be brought to health care, where in many cases the current system makes it easier for an employee to make a mistake than it is not to make a mistake (drugs with similar names are a common example). The more rapidly and completely the health care industry is able to fully integrate this type of systems-based thinking and escape the traditional paradigm of blame, the faster we will be able to make real improvements in patient care.

High-Level Leadership. A common theme in many quality improvement efforts is the need for strong leadership from the top of an organization. In an auto company, this means that the top executives are all fully committed to the quality improvement effort. Similarly, in health care the top executives in a hospital must be totally committed to a culture of improvement. However, health care has an added layer of complexity compared to industry. In health care, each physician is in many respects a top leader. Because of this, it is important to have the buy-in of the physicians in an organization if a full commitment to quality improvement is to be made. All quality improvement efforts, but particularly methods as culturally-engrained as Six Sigma and TPS, require a substantial investment of staff time, money, and institutional resources.

In tight budget times, this investment may be difficult for hospital executives to justify, but it is critical to the advancement of quality improvement efforts. Fortunately, the need for a systemic commitment is beginning to be appreciated and advanced by clinical leaders.[84] It is imperative that this type of support continue and even be increased in order to accelerate the progress of quality improvement.

Financial Incentives. Perhaps the greatest challenge to quality improvement is the fact that our current medical reimbursement system is not designed to pay for high quality care. Although it is well established that, in general, high quality care costs less in the long run than low quality care, the investments and rewards for quality of care are often not aligned.[8] Paying for quality will require a major redesign of the business models for health care organizations. Our payment system is still fundamentally a fee-for-service system. This system can create "perverse behaviors" in which the primary incentive is to provide more care rather than better care.[85] At the current time, we have no good way to pay for preventive care, such as paying physicians for *preventing* cases of blindness among diabetic patients. Similarly, we have no way of paying hospitals for *prevented* cases of central line infections or pressure ulcers. Until major changes to our payment structure take place, health care organizations will be limited in the resources they can devote to quality improvement.

Summary.

Evidence-based medicine is today the standard for clinical care. New interventions are not embraced widely until their safety, efficacy, and effectiveness have been demonstrated through methodologically rigorous studies. This same expectation of evidence has not yet fully filtered to the field of health care quality improvement, which continues to share strong ties with the industrial improvement world from which it developed. The emphasis of much quality improvement is on the pragmatic concerns of achieving a workable improvement in a process as quickly and inexpensively as possible. In many industries, this is a perfectly reasonable approach. However, if quality improvement is to become an integral part of clinical care, its methods must be vetted in a way with which clinicians are comfortable.

Six Sigma and TPS are examples of successful business models for improvement that are beginning to be used in health care. The use of these systems is limited by a lack of adequate evidence that supports their application to health care processes. Success of these systems demands the support of clinical staff, and these clinical staff demand that adequate literature exists to justify their use.

In this review, I have laid out basic principles to which quality improvement projects should conform if they are destined for publication and designed to be used to convince clinicians to embrace quality improvement methods. Projects should be clinically relevant, they should utilize a conceptually solid improvement plan such as the PDSA cycle, they should be designed with adequate control and comparison groups to make the conclusions meaningful, and

the results should be reported with enough methodological and statistical detail to enable clinicians to feel confident about the methods and results.

Routinely carrying out improvement projects that meet the above requirements will require a commitment from the institution, the quality improvement professional, and the clinician. This level of commitment is made worthwhile by the potential that quality improvement methods have to make real progress in narrowing the chasm between our current health care delivery system and our vision of the ideal delivery system.

Conclusion

Like Abraham Flexner's report on medical education in 1910, the Institute of Medicine reports nearly a century later have set in motion a dramatic change in the U.S. health care landscape. In *To Err is Human*, the IOM called attention to the dramatic mortality statistics attributable to medical error. In *Crossing the Quality Chasm*, they laid out the destination for which American health care should aim. It has been left to others to figure out how to span the substantial chasm between where we are today and where we want to be. Perhaps uncertainty about how to make real progress toward the lofty goals of the IOM explains the relative lack of progress in achieving the goals in the five years since publication of *Crossing the Quality Chasm*.[86]

Quality improvement has been a formal part of some fields of health care for over half a century. However, it has been only much more recently that quality improvement has begun to be viewed as integral to every aspect of health care. In response to this change in prominence, many individuals and

organizations have begun to search out methods to help make the kinds of dramatic shifts in our system of health care that are needed in order to reach the goals laid out by the IOM. Two methods that have recently been embraced by the health care industry are Six Sigma and Toyota Production System.

Both Six Sigma and TPS were developed in industry, where their success is beyond question. Their utility and potential for success in medicine is not as certain. Despite being embraced by numerous health care organizations, little objective evidence exists to demonstrate that these methods are appropriate to the world of patient care. This systematic review found only five published studies that met even the most basic criteria for being considered objective measurements of the use of Six Sigma or TPS in health care.

The great paradox of Six Sigma and TPS is that both require total buy-in from all members of an organization if they are to be successful, but evidence to encourage that buy-in in health care systems is lacking. Without clinician buy-in to these methods, they are unlikely to produce results in health care. At the same time, clinicians are understandably hesitant to make all-out commitments to the methods of Six Sigma and TPS without evidence of effectiveness that meets the standards clinicians have come to expect from biomedical literature.

The solution to this situation is clear but certainly not simple. Quality improvement projects that utilize Six Sigma, TPS, and other industry-derived methods must be conducted with sufficient rigor to allow for publication of credible studies that can convince clinicians of the value of the method. This shift from designing quality improvement studies with only pragmatic concerns to

designing them with methodological concerns will require significant investment on the part of institutions and clinicians. Such investment may often be elusive in the busy, financially difficult world of health care.

Quality improvement in health care is here to stay, and the magnitude of the changes needed in the health care system is such that real progress requires widespread adoption of an accepted, validated tool set. Six Sigma and TPS may be able to provide frameworks for such a tool set, but their application in health care will continue to be very limited until sufficient high-quality evidence to their effectiveness can be amassed in the literature.

References

1. Kohn LT, Corrigan J, Donaldson MS. *To err is human : building a safer health system*. Washington, D.C.: National Academy Press; 2000.
2. Brennan TA. The Institute of Medicine report on medical errors--could it do harm? *Tex Med.* Jun 2000;96(6):13-15.
3. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. *Jama.* Jul 5 2000;284(1):93-95.
4. Leape LL. Institute of Medicine medical error figures are not exaggerated. *Jama.* Jul 5 2000;284(1):95-97.
5. Institute of Medicine (U.S.). Committee on Quality of Health Care in America. *Crossing the quality chasm : a new health system for the 21st century*. Washington, D.C.: National Academy Press; 2001.
6. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. *Jama.* Sep 16 1998;280(11):1000-1005.
7. Urbina I. In the treatment of diabetes, success often does not pay. *The New York Times.* January 11, 2006, 2006.
8. Leatherman S, Berwick D, Iles D, et al. The business case for quality: case studies and an analysis. *Health Aff (Millwood).* Mar-Apr 2003;22(2):17-30.

9. Flexner A. Medical education in the United States and Canada. From the Carnegie Foundation for the Advancement of Teaching, Bulletin Number Four, 1910. *Bull World Health Organ.* 2002;80(7):594-602.

10. Flexner A, Vevier C, University of Medicine and Dentistry of New Jersey. *Flexner : 75 years later : a current commentary on medical education.* Lanham, MD: University Press of America; 1987.

11. Braunwald E. *Harrison's principles of internal medicine.* 15th ed. New York: McGraw-Hill; 2001.

12. Doherty S. History of evidence-based medicine. Oranges, chloride of lime and leeches: barriers to teaching old dogs new tricks. *Emerg Med Australas.* Aug 2005;17(4):314-321.

13. Donabedian A. *The criteria and standards of quality.* Ann Arbor, Mich.: Health Administration Press; 1982.

14. Larson JS, Muller A. Managing the quality of health care. *J Health Hum Serv Adm.* Winter 2002;25(3):261-280.

15. Luce JM, Bindman AB, Lee PR. A brief history of health care quality assessment and improvement in the United States. *West J Med.* Mar 1994;160(3):263-268.

16. Langley GJ. *The improvement guide : a practical approach to enhancing organizational performance.* 1st ed. San Francisco: Jossey-Bass Publishers; 1996.

17. Colton D. Quality improvement in health care. Conceptual and historical foundations. *Eval Health Prof.* Mar 2000;23(1):7-42.

18. Harvey G. Quality in health care: traditions, influences and future directions. *Int J Qual Health Care.* Aug 1996;8(4):341-350.

19. Deming WE. *The new economics for industry, government, education.* 2nd ed. Cambridge, MA: Massachusetts Institute of Technology, Ctr. for Adv. Eng. Study; 1994.

20. Berwick D, Kabcenell A, Nolan T. No Toyota yet, but a start. A cadre of providers seeks to transform an inefficient industry--before it's too late. *Mod Healthc.* Jan 31 2005;35(5):18-19.

21. Lasky FD, Boser RB. Designing in quality through design control: a manufacturer's perspective. *Clin Chem.* May 1997;43(5):866-872.

22. Motorola. Motorola University Six Sigma Dictionary. http://www.motorola.com/content.jsp?globalObjectId=3074-5804. Accessed April 27, 2006.

23. Scalise D. Six sigma. The quest for quality. *Hosp Health Netw.* Dec 2001;75(12):41-46, 42.

24. Goedert J. Crunching data: the key to six sigma success. *Health Data Manag.* Apr 2004;12(4):44-48.

25. Hagland M. Six sigma: it's real, it's datadriven, and it's here. *Health Care Strateg Manage.* Dec 2005;23(12):1, 13-16.

26. Motorola. About Motorola University. http://www.motorola.com/content.jsp?globalObjectId=3071-5801. Accessed April 27, 2006.

27. *General Electric Company 1997 Annual Report.* Fairfield, CT 1998 1997.

28. Elgert S. Quality improvement in family practice: a recipe for change. *Am J Med Qual.* Mar-Apr 2004;19(2):83-87.

29. Spear SJ. Learning to lead at Toyota. *Harv Bus Rev.* May 2004;82(5):78-86, 151.

30. Appleby C. Industrial strength. Patient safety means leading pre-industrial health care into a post-industrial world. *Trustee.* Jan 2002;55(1):10-14, 19, 12.

31. Liker JK. *The Toyota way : 14 management principles from the world's greatest manufacturer*. New York: McGraw-Hill; 2004.

32. Monden Y. *Toyota production system : an integrated approach to just-in-time*. 2nd ed. Norcross, Ga.: Industrial Engineering and Management Press; 1993.

33. Johnstone PA, Hendrickson JA, Dernbach AJ, et al. Ancillary services in the health care industry: is Six Sigma reasonable? *Qual Manag Health Care.* Jan-Mar 2003;12(1):53-63.

34. Bartlett RC. Trends in quality management. *Arch Pathol Lab Med.* Nov 1990;114(11):1126-1130.

35. Kazmierczak SC. Laboratory quality control: using patient data to assess analytical performance. *Clin Chem Lab Med.* May 2003;41(5):617-627.

36. Stankovic AK. The laboratory is a key partner in assuring patient safety. *Clin Lab Med.* Dec 2004;24(4):1023-1035.

37. Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. *Arch Pathol Lab Med.* Apr 2000;124(4):516-519.

38. GE Healthcare- Product Features- Performance Improvement. http://www.gehealthcare.com/usen/service/products/performance.html. Accessed May 2, 2006.

39. Benedetto AR. Adapting manufacturing-based Six Sigma methodology to the service environment of a radiology film library. *J Healthc Manag.* Jul-Aug 2003;48(4):263-280.

40. Benedetto AR, Dunnington JS, Oxford-Zelenske D. Using Six Sigma to improve the film library. *Radiol Manage.* Sep-Oct 2002;24(5):26-35; quiz 36-28.

41. Loree A, Maihack M, Powell M. The path of least resistance: is there a better route? *Radiol Manage.* Sep-Oct 2003;25(5):48-51.

42. Higgins TJ, Diener JC, Grunsfeld T, Wogen S. Aim for perfection. Evaluate e-prescribing technology using six-sigma techniques. *MGMA Connex.* Oct 2002;2(9):35-38.

43. Young D. Six Sigma black-belt pharmacist improves patient safety. *Am J Health Syst Pharm.* Oct 1 2004;61(19):1988, 1992, 1996.

44. Revere L, Black K. Integrating Six Sigma with total quality management: a case example for measuring medication errors. *J Healthc Manag.* Nov-Dec 2003;48(6):377-391; discussion 392.

45. Rosen L. Applying industrial engineering practices to radiology. *Radiol Manage.* Nov-Dec 2004;26(6):32-35.

46. Thompson DN, Wolf GA, Spear SJ. Driving improvement in patient care: lessons from Toyota. *J Nurs Adm.* Nov 2003;33(11):585-595.

47. Berlowitz DR. Striving for six sigma in pressure ulcer care. *J Am Geriatr Soc.* Sep 2003;51(9):1320-1321.

48. Simmons JC. Using Six Sigma to make a difference in health care quality. *Qual Lett Healthc Lead.* Apr 2002;14(4):2-10, 11.

49. Morgan SP, Cooper C. Shoulder work intensity with Six Sigma. *Nurs Manage.* Mar 2004;35(3):29-32.

50. Make these changes to cut delays, diversion hours. *ED Manag.* Jun 2003;15(6):64-65.

51. Ettinger WH. Six Sigma: adapting GE's lessons to health care. *Trustee.* Sep 2001;54(8):10-15, 11.

52. Scalise D. Six sigma in action. Case studies in quality put theory into practice. *Hosp Health Netw.* May 2003;77(5):57-62, 52.

53. Sherman J. Achieving real results with Six Sigma. "Six Sigma to the rescue," declared the title of a June 2002 article in the technology section of Health Care Finance. Almost four years later, has Six Sigma helped healthcare organizations achieve the promised breakthrough improvement in their operations? *Healthc Exec.* Jan-Feb 2006;21(1):8-10, 12-14.

54. Spear SJ. Fixing health care from the inside, today. *Harv Bus Rev.* Sep 2005;83(9):78-91, 158.

55. Stymiest DL. Analyze this! Applying Six Sigma concepts to facilities maintenance. *Health Facil Manage.* Jun 2005;18(6):21-24.

56. Leape LL, Bates DW, Cullen DJ, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA.* Jul 5 1995;274(1):35-43.

57. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med.* Jun 26 2003;348(26):2635-2645.

58. Chan AL. Use of Six Sigma to improve pharmacist dispensing errors at an outpatient clinic. *Am J Med Qual.* May-Jun 2004;19(3):128-131.

59. Kang JO, Kim MH, Hong SE, Jung JH, Song MJ. The application of the Six Sigma program for the quality management of the PACS. *AJR Am J Roentgenol.* Nov 2005;185(5):1361-1365.

60. Adams R, Warner P, Hubbard B, Goulding T. Decreasing turnaround time between general surgery cases: a six sigma initiative. *J Nurs Adm.* Mar 2004;34(3):140-148.

61. Frankel HL, Crede WB, Topal JE, Roumanis SA, Devlin MW, Foley AB. Use of corporate Six Sigma performance-improvement strategies to reduce incidence of catheter-related bloodstream infections in a surgical ICU. *J Am Coll Surg.* Sep 2005;201(3):349-358.

62. Raab SS, Andrew-Jaja C, Condel JL, Dabbs DJ. Improving Papanicolaou test quality and reducing medical errors by using Toyota production system methods. *Am J Obstet Gynecol.* Jan 2006;194(1):57-64.

63. Benedetto AR. Six Sigma: not for the faint of heart. *Radiol Manage.* Mar-Apr 2003;25(2):40-53.

64. Staker LV. Teaching performance improvement: an opportunity for continuing medical education. *J Contin Educ Health Prof.* Spring 2003;23 Suppl 1:S34-52.

65. Ovretveit J. Would it work for us? Learning from quality improvement in Europe and beyond. *Jt Comm J Qual Improv.* Jan 1997;23(1):7-22.

66. Rubenstein LV, Fink A, Yano EM, Simon B, Chernof B, Robbins AS. Increasing the impact of quality improvement on health: an expert panel method for setting institutional priorities. *Jt Comm J Qual Improv.* Aug 1995;21(8):420-432.

67. Ettinger W, Van Kooy M. The art and science of winning physician support for Six Sigma change. *Physician Exec.* Sep-Oct 2003;29(5):34-38.

68. Speroff T, James BC, Nelson EC, Headrick LA, Brommels M. Guidelines for appraisal and publication of PDSA quality improvement. *Qual Manag Health Care.* Jan-Mar 2004;13(1):33-39.

69. Speroff T, O'Connor GT. Study designs for PDSA quality improvement research. *Qual Manag Health Care.* Jan-Mar 2004;13(1):17-32.

70. Kizer KW. The "new VA": a national laboratory for health care quality management. *Am J Med Qual.* Jan-Feb 1999;14(1):3-20.

71. McQueen L, Mittman BS, Demakis JG. Overview of the Veterans Health Administration (VHA) Quality Enhancement Research Initiative (QUERI). *J Am Med Inform Assoc.* Sep-Oct 2004;11(5):339-343.

72. Feinstein KW, Grunden N, Harrison EI. A region addresses patient safety. *Am J Infect Control.* Jun 2002;30(4):248-251.

73. *The Breakthough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement.* Boston: Institute for Healthcare Improvement; 2003.

74. Wholey JS, Hatry HP, Newcomer KE. *Handbook of practical program evaluation.* 2nd ed. San Francisco: Jossey-Bass; 2004.

75. Henry SB. Informatics: essential infrastructure for quality assessment and improvement in nursing. *J Am Med Inform Assoc.* May-Jun 1995;2(3):169-182.

76. Davidoff F. Databases in the next millennium. *Ann Intern Med.* Oct 15 1997;127(8 Pt 2):770-774.

77. Davies HT. Exploring the pathology of quality failings: measuring quality is not the problem--changing it is. *J Eval Clin Pract.* May 2001;7(2):243-251.

78. Carey RG, Lloyd RC. *Measuring quality improvement in healthcare : a guide to statistical process control applications.* New York: Quality Resources; 1995.

79. Plsek PE. Quality improvement methods in clinical medicine. *Pediatrics.* Jan 1999;103(1 Suppl E):203-214.

80. James C. Manufacturing's prescription for improving healthcare quality. *Hosp Top.* Winter 2005;83(1):2-8.

81. Levett JM, Carey RG. Measuring for improvement: from Toyota to thoracic surgery. *Ann Thorac Surg.* Aug 1999;68(2):353-358; discussion 374-356.

82. Berte LM. Patient safety: getting there from here--quality management is the best patient safety program. *Clin Leadersh Manag Rev.* Nov-Dec 2004;18(6):311-315.

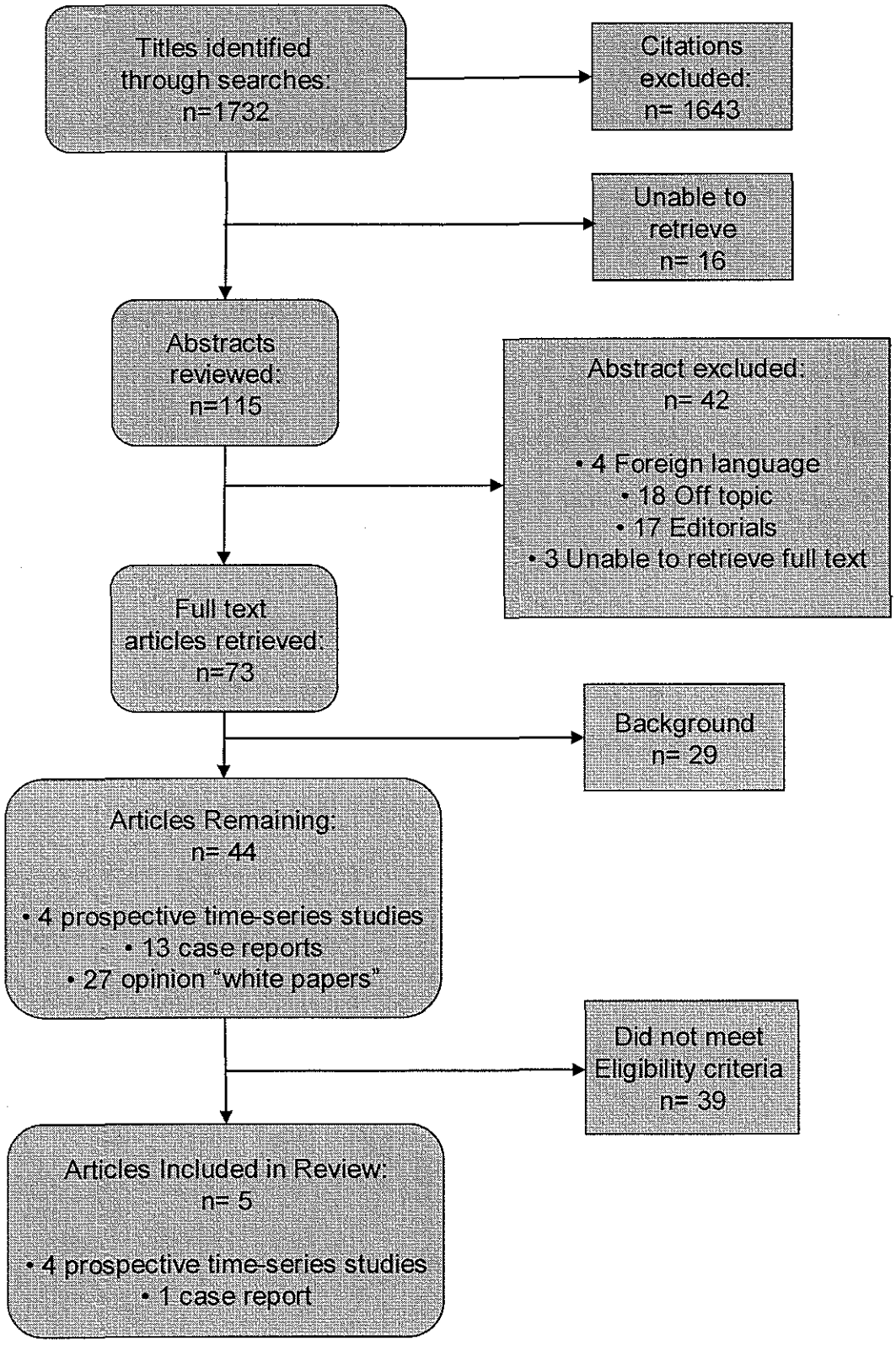
83. Chalice R. *Stop rising healthcare costs using Toyota lean production methods : 38 steps for improvement.* Milwaukee, Wis.: ASQ Quality Press; 2005.

84. Moorman DW. On the quest for Six Sigma. *Am J Surg.* Mar 2005;189(3):253-258.

85. Coye MJ. No Toyotas in health care: why medical care has not evolved to meet patients' needs. *Health Aff (Millwood).* Nov-Dec 2001;20(6):44-56.

86. Longo DR, Hewett JE, Ge B, Schubert S. The long road to patient safety: a status report on patient safety systems. *JAMA.* Dec 14 2005;294(22):2858-2865.

Appendix I: QUORUM Tree of Literature Search



Appendix II: Evidence Table for Studies Meeting Inclusion Criteria

Author, Study	Design	Outcome	Quality	Generalizability
Chan AL. Use of Six Sigma to improve pharmacist dispensing errors at an outpatient clinic	Time series/case report	Decreased dispensing error rate from 338.8 DPMO to 230 DPMO	Poor	Unable to Assess
Kang JO. The application of the Six Sigma program for the quality management of the PACS	Prospective time series	Decreased maintenance requirements to 79% of pre-intervention level	Fair/ poor	Moderate, given the standardization of PACS systems
Adams R. Decreasing turnaround time between general surgery cases: a six sigma initiative	Prospective time series	Decreased Z score for cases violating upper control limit from 2.13 to 1.53	Fair	Moderate, given commonality of OR turnaround time problems and processes
Frankel HL. Use of corporate Six Sigma performance-improvement strategies to reduce incidence of catheter-related bloodstream infections in a surgical ICU	Prospective time series	CR-BSI rate decreased from 11/1000 catheter days to 1.7/1000 catheter days ($p<0.0001$)	Fair	Good for SICUs with similar patient and staffing profiles
Raab SS. Improving Papanicolaou test quality and reducing medical errors by using Toyota production system methods	Noncon-current cohort study	Frequency of absent transformation zone cells decreased from 9.86% to 4.74% ($p=0.001$) Percentage of tests diagnosed as ASC-US decreased from 7.8% to 3.9% ($p=0.008$)	Poor	Unable to assess