

No Respect: Research in Quality, Safety, and Process Improvement

Ilan S Rubinfeld, MD, MBA
H Mathilda Horst, MD

Abstract

The need for good quality and safety research has never been more imperative, but even as we encourage and promote such work, we seem to suppress it through institutional bias and inertia. Indeed the culture of health care seems to have a love-hate relationship with quality-improvement work as a whole. In this commentary we explore some of the implications of the application of pure science standards at the sharp end of clinical practice, where the down-and-dirty street-level improvement work happens.

The realm of biomedical publications is an ever-expanding multidisciplinary collection of books, journals, Internet and multimedia resources. The diversity of material and types of authors in that realm has exploded. Concurrently, many of the old guard would say that the number of high-quality projects has diminished commensurate with the increase in competition for dwindling research funds and the widening gap favoring bioscientific staff over physician researchers in awards for scarce National Institutes of Health resources. Additionally, newer kinds of research, mature and well developed in other industries, are now emerging in health care to meet the burning needs associated with behavioral and process issues in our industry.

Among previous generations, even in the 1990s, current successes in quality health care would not have been well received. It was not previously acceptable to quote Edward Deming, MD, to a medical audience, nor could we have elevated individuals like Donald Berwick, MD, to rock-star status in the 1970s unless his work had revolved around laboratory rats or resulted in naming a few syndromes. Hospital chief executive officers were not then coveting a Malcolm Baldrige National Quality Program award, creating budgets for

the 5 Million Lives Campaign, or planning to eliminate harm of all sorts and “never events.” Yet we still hold on to the view of research and scientific inquiry as being traditional: the necessary and only means by which we can effect change. This traditional perspective positions level 1 data generated from prospective, randomized, controlled, blinded, multicenter trials (RCTs) as the holy grail of all research. Indeed, within our institution’s training programs, we teach this same gospel and analyze our journal club articles applying the levels of evidence-based medicine. We can recall the fateful day we heard Terry P Clemmer, MD, speak at the Institute for Healthcare Improvement (IHI)¹ about “pragmatic research” and the repugnance it inspired in us; he is now a visionary in our view.

This commentary is meant not to criticize traditional scientific inquiry, its purity or its capability of answering our questions; it is meant to ponder the new question: how do we look at real-world, problem-solving solutions in the realm of process and performance improvement using newer methods more appropriate to evaluate these activities?

“I don’t get no respect,”
—Rodney Dangerfield

So much of the improvement literature demonstrates the necessity to solve problems based on local needs and operational issues, a concept that lessens the effectiveness of a multicenter trial. Indeed, many brilliantly conducted prospective trials have limited effectiveness once study conditions are no longer present and findings are implemented in diverse and disparate environments. A tradition and track record for operational effectiveness work is that of the factory where statistical tolerance and variability is monitored closely and improved based on historical local controls.

Ilan S Rubinfeld, MD, MBA, is an Attending Physician in Trauma and Critical Care and Acute Care Surgery at Henry Ford Hospital in Detroit, MI. He is also Assistant Professor of Surgery, Wayne State University School of Medicine; Associate Program Director, General Surgery Residency, Henry Ford Hospital; and Associate Medical Director, Surgical Intensive Care, Henry Ford Hospital. E-mail: irubinfeld@yahoo.com.

H Mathilda Horst, MD, is a Surgeon at the Henry Ford Hospital in Detroit, MI. E-mail: mhorst1@hfhs.org.

This is automatically second- and third-tier literature for health care. It may be responsible for explaining the market decline of the American automobile industry and worth literally billions of dollars to companies such as Honda and Toyota for focusing on real-world improvements, quality, and customer satisfaction. As we pursue ideals of our medical product being reproducible, reliable, and excellent, is there nothing we can learn from such success?

Under a medical journal's traditional editorial review process, Toyota would have faced stark criticism for not randomizing its factories to lean and not lean in some way. Maybe they could have been lean by the week or month and tracked the vehicle identification numbers? Maybe automakers could have made Toyota about customer dissatisfaction and Lexus about satisfaction for better control? Perhaps they could have manufactured a few random vehicles in Detroit to see if quality and satisfaction were maintained. Or, like a classic health care stance, they could believe that their vehicles are more complicated and harder to manufacture than the vehicles of others. Perhaps these examples seem irrelevant compared with scientific research. Perhaps typical medical readers think that we are far from any medical relevance to their practice or clinical realm. Nothing could be further from the truth as long as at the greatest of US hospitals, individual patients suffer unacceptable outcomes.

This invited commentary came in response to our experience in submitting a paper to this journal, in going through the process of receiving reviewer comments and in our responses to them. We knew our manuscript didn't represent a defining moment in clinical quality research. The article simply described our improvements in the realm of tight glycemic control in the population of patients undergoing bariatric surgery.¹ After a broad flurry throughout the literature of initial publications regarding individual institutions' successes and problems, we pursued our project as a second-generation one. Specifically, our article addressed how we adapted our institutional protocol to a high-risk group that did not apparently benefit from the broader protocol. For those of us working in multidisciplinary teams trying to make sense of the literature, the logistics of these protocols and how to apply them to our various populations, we thought our approach would be valuable to others asking similar questions. We wished someone had written such an article for us two years earlier to share their experience and offer guidance, so we wrote the article with the intention of sharing the results of our pain and suffering. We chose *The Permanente Journal*

because of its intention to publish articles regarding quality and outcomes.

Our manuscript received comments by four reviewers, all of whom provided detailed questions and concerns. Along the review-revision pathway, we noticed that the reviewers' questions displayed the weakness of the traditional-research view in this pragmatic research literature realm. This is not a personal criticism of our reviewers or the community of reviewers who do outstanding, thankless volunteer work with a high level of investment and commitment. We raise the issue of whether the traditional review parameters and traditional-minded reviewers can serve this type of research's intention, as noted in the report by Davidoff and Batalden² advocating quality-improvement publication guidelines.

Most process-improvement research is done by evaluating existing processes and changes implemented for improvement. The artifacts and evidence of process tend to be found in various data sets available to the committees and teams that work on these issues. Much like tuning up the plant floor operations in the automotive industry, similar statistics, metrics, and chosen performance improvement benchmarks are monitored. Tools such as clinical dashboards are routinely reviewed with the intent to make improvements in patient care and outcomes. Certain outcomes may be process outcomes whereas others are the more direct objective related to actual patient outcomes. For example, a group seeking to improve ventilator-associated pneumonia (VAP) may choose to implement a VAP treatment protocol bundle. As a process measure, bundle compliance may be monitored as a whole or by bundle component. In the end, group members are most interested in decreasing measured rates of VAP, but they must assess their process to see how they are doing before claiming success or rejecting the current mechanism. These clinical dashboards are critical to the ongoing logistics and maintenance of hospital care, yet are apparently weak and useless to the purest scientific researcher.

In traditional clinical research, the inclusion and exclusion criteria are critical. A savvy reviewer can often uncover clear errors or weaknesses in the methodology by careful focus. For our bariatric article, several reviewer comments were focused on this topic. Indeed, in anticipation of this focus we carefully and clearly detailed our inclusion criterion. In our study *all* patients were included. *All* of the bariatric patients were also

[In anticipation of reviewing comments] ... we carefully and clearly detailed our inclusion criterion. In our study *all* patients were included.

included. The comparison came from those bariatric patients included in *all* before the targeted change and those patients included in *all* after the change. This disconnect was based on philosophical differences in process. In the process and safety realm, it is preferable to treat as many populations in as similar a way as possible to eliminate unnecessary complexity. Exceptions are made based on data, but not without some substantiation.

Similarly, administrative and quality data are observational and retrospective in nature, unlike a prospective trial where a data-collection sheet is created ahead of time and with any luck anticipated variables are collected prospectively. In the world of quality improvement as a day-by-day struggle, we often need to be opportunistic and pragmatic using existing and limited data as best we can. Indeed, we are not even usually specifically engaged in the explicit pursuit of research as much as we are engaged in the care process, trying to fix what ails our systems of care and to uncover what risks and harm our patients face. Prospective trials exist in quality, though they are rare and difficult to accomplish. The Keystone collaborative³ is such a process, and we are excited and engaged participants in this method. Neither approach is better or best: they are different and must be judged using different tools.

Both perspectives are iterative. Purely scientific researchers do increasingly complex and subspecialty work on new concepts they have developed. They describe a scientific observation, validate it, then work to deepen the understanding and limits of knowledge. This process works for choosing the right operation, deciding which antibiotic to give for a certain infection, or documenting the effects on cell culture of a novel enzyme or viral probe. In process improvement the iterative nature is developed through improvement cycles called total quality or plan-do-check-act cycles, which set up the iterative process as a multiyear effort to improve. If made to meet the burdens of traditional scientific research, each iteration would be rejected soundly. Similarly, the incredible results attributed to rapid-cycle improvement both in health care and other industries would not have occurred in the environment of rigid controls necessary for good traditional science.

As already mentioned, several years ago we were fortunate to hear a presentation by Dr Terry Clemmer at a national IHI conference.⁴ He presented an introduction to pragmatic research. We sat in anger and frustration, alarmed by his words, fiercely clinging to our then traditional view of medical research. In academic surgery circles the traditional “success” defining

a physician combines a great clinician, teacher, and laboratory researcher. To be successful in all three was necessary for advancement and recognition, and the expectation existed that all faculty would want to be a true “triple threat.” Clinical articles have always lacked the prestige of excellent basic science, with the rare exception of RCTs, of which there are relatively few. In this reality, the quality and safety enthusiast may be viewed as a scoundrel of little merit. The chief quality officer was often symbolic rather than someone with teeth and real influence. Perhaps with the now ever-present push for real quality and outcomes data that are based on acuity adjustment and the need for transparency for our communities and patients, we can begin to treat process work as its own important world in need of understanding, nurturing, and development.

Our institution and our Department of Surgery have taken on the National Surgical Quality Improvement Program (NSQIP) processes as a way of life and operational improvement. Patients are sampled and their outcomes extrapolated. Yet interventions to improve our data are based on the entire population. The NSQIP, as a nationally accepted and celebrated process, has created the opportunity to use acuity-adjusted outcomes data to compare between hospitals and level-set expectations. This represents a key validation that process is important and does affect outcomes. In practice, the majority of research productivity from NSQIP relates to perioperative care issues.

Even our own institutional review board (IRB) has difficulty with process work. The traditional IRB is staffed by basic scientists and many nonclinicians. To an immunologist or geneticist, pragmatic research is inelegant. Emphasis on sample size and research design are critical in a front-loaded process such as a prospective trial, where all contingencies must be analyzed before exposing patients to such risk or the purchase of expensive reagents. In quality process research we can pull seemingly endless amounts of data if necessary, but we may not always find what we want. We can go back to the computer and pull more variables if we like. We might generate more questions as we go and incur no added risk by re-exploring the data sources. We may be put down as “data miners” and as engaging in “hypothesis-generating research.” It seems that only in the nonmedical halls of academia, the business school, and the occasional engineering doctoral candidate can appreciate a well-executed rapid cycle improvement.

Yet the future is rosy, and progress can be made. As recently as 2003, at our departmental grand rounds, a

In the world of quality improvement as a day-by-day struggle, we often need to be opportunistic and pragmatic using existing and limited data as best we can.

brave resident presented some hard-won data on our tight glycemic protocol in cardiothoracic patients. The carnage was memorable and we as his faculty advisors were similarly gored. Yet, in the spring of 2008, during the annual research presentations of our residents, more than six projects were presented in the realm of quality and process improvement. No blood fell on the floor; the residents and faculty were engaged and full of questions as they saw the relevance and importance to their patients' care. Since 2006, we have conducted a quarterly NSQIP reporting session and operations meeting. We have plans to pursue multispecialty status with NSQIP. Pragmatic research is a street-level bedside struggle to improve patient care with tools that focus on people, behavior, and process. It is a long-term, team struggle in the trenches with nurses, physicians, pharmacists, respiratory therapists, and many others. It is a soccer game of continuous action rather than a football game's brief glorious moments of action. It is the long-shunned relative of traditional bench research, worthy of support and nurturing.

Pragmatic research should be judged and held to high standards like those recommended by Davidoff and Batalden,² Berwick,⁴ and Thomson⁵ in 2005, using standards different from those used to analyze traditional bench research. We cannot use a German test to evaluate a student studying Spanish. Evaluation methods used for traditional scientific studies do not serve the root cause of quality-improvement studies; square pegs will not fit round holes. Dr Berwick⁶ illustrated this evaluation chasm in what has occurred with rapid-response team outcomes reported by individual hospitals (beneficial) versus a cluster RCT (nonbeneficial), creating a continuing controversy over the scientific worth of these teams. We favor Dr Berwick's argument that quality-improvement initiatives, such as

rapid-response teams, are "a process of social change," requiring changes in the evaluation approach for these reports. We call upon our industry's wise journal editors and reviewers to take heed of the differences between traditional science and the science of improvement to enact long-needed change in the evaluation methods for quality-improvement publications. ❖

^a Implementing an idealized model for critical care. Workshop. 18th Annual National Forum on Quality Improvement in Health Care. Orlando, FL: 2006 Oct 10-13.

Disclosure Statement

The author(s) have no conflicts of interest to disclose.

Acknowledgment

Katharine O'Moore-Klopf, ELS, of KOK Edit provided editorial assistance.

References

1. Joseph B, Genaw J, Carlin A, Jordan J, Talley J, Rubinfeld I. Perioperative tight glycemic control: the challenge of bariatric surgery patients and the fear of hypoglycemia. *Perm J* 2007 Spring;11(2):36-9.
2. Davidoff F, Baltaden P. Toward stranger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005 Oct;14(5):319-25.
3. Michigan Health and Hospital Association Keystone Center for Patient Safety and Quality [Web page on the Internet]. Lansing, MI: Michigan Health and Hospital Association; 2003-2008 [cited 2008 Feb 6]. Available from: www.mhakeystonecenter.org.
4. Berwick DM. Broadening the view of evidence-based medicine. *Qual Saf Health Care* 2005 Oct;14(5):315-6.
5. Thomson RG. Consensus publication guidelines: The next step in the science of quality improvement? *Qual Saf Health Care* 2005 Oct;14(5):317-8.
6. Berwick DM. The science of improvement. *JAMA* 2008 Mar 12;299(10):1182-4.

Question

In research, and probably also in practice,
maintaining and fostering curiosity—
the ability to ask questions each time a new
phenomenon occurs is indispensable.

— Baruch S Blumberg, b 1925, 1976 recipient of the Nobel Prize in Physiology and Medicine