The poverty of theory: Evidence-based medicine and the social contract

Steven R. Shackford, MD, Coronado, California

In his book entitled, “The Poverty of Theory and Other Essays,” British historian, E.P. Thompson, placed a quote by Karl Marx on the frontispiece: “To leave error unrefuted is to encourage intellectual immorality.” As surgeons, we do our best to correct errors—we carefully scrutinize and analyze errors in the morbidity and mortality review, which has become an icon in the lore of surgery. The question is does medicine, as a profession, make the same effort as we surgeons to refute error? If not, are we failing to leave error unrefuted, thus eroding the public trust and failing in our duty to comply with the social contract? Sadly, I believe that the answer to all three questions is affirmative, and I will use quality in health care as the platform for demonstrating this. In this context, I will review the nature and importance of the social contract. I will provide a truncated review of the history of the health care quality movement and, hopefully, demonstrate the critical need for evidence-based medicine. Finally, I am going to review two examples of unrefuted error that were related to quality and had some significant unintended consequences. My purpose, like the purpose of our morbidity and mortality conference, is not to critique the past, but to prevent repetition in the future.

Do you ever wonder why it is, when you come into an examining room to see a new patient for the first time, that those patients, within minutes, are willing to tell you the deepest, darkest secrets of their lives, take off all of their clothing, and allow you to feel and probe? It is not because they respect you. It is because of what thousands of physicians, nurses, and scientists who have gone before you have accomplished and made medicine what it is today, a highly respected and highly trusted profession. It is, therefore, incumbent upon all of us, as a fiduciary responsibility, to continue to build on the respect of the profession.

Our responsibility to our patients and the public’s reciprocal trust that we will always act in their interest forms the basis of medicine’s contract with society. Social contracts have existed since ancient Rome. In essence, society surrenders a freedom for protection of other freedoms. For us, society gives our profession autonomy in the affairs of medicine. Autonomy means that we certify, we verify, and we test our peers for competency.

Medicine reciprocates by providing high-quality care, standards of ethics, self-regulation, and affirms the validity of the science underlying medical practice. We act to further the evidence base that supports our decision making and we rid the profession of incompetent practitioners. In short, “Society granted physicians status, respect, autonomy in practice, the privilege of self-regulation, and financial rewards on the expectation that physicians would be competent, altruistic, moral, and would address the health care needs of individuals and society. This “arrangement” remains the essence of the social contract.” This is the normative behavior of physicians. It is the public expectation. Occasionally, some physicians violate that contract. I will highlight two such violations.

You all are familiar with gabapentin (Neurontin). It was patented in 1977 and approved by the Federal Drug Administration (FDA) for the treatment of complex seizures in 1993. David Franklin, a PhD pharmacist, working for Pfizer (the manufacturer of gabapentin) discovered that the company was advertising off-label uses, which is prohibited by the FDA. Pfizer was also hiring academic physicians to advocate for off-label use, in violation of FDA principles and prohibitions. Some of these individuals were receiving up to US $40,000 a year for this advocacy. In 2004, Warner-Lambert (parent of Pfizer) agreed to plead guilty and to pay more than US $430 million to resolve criminal charges and civil liabilities. The “Neurontin legacy” has led some to suggest that drastic action is essential to preserve the integrity of medical science and practice and to justify the public trust.

The second example concerns outright physician fraud that was featured in most of the nation’s newspapers with this headline, “412 Physicians Charged in a US $1.3 Billion Opioid Fraud Takedown.” Our Attorney General was quoted, “Too many trusted medical professionals have chosen to violate their oaths and put greed ahead of their patients.” I am sure that all of you know of other examples, but I think I have made my point.

To make my case regarding unrefuted errors, some historical background is necessary. Quality in health care began with Florence Nightingale, who I consider to be the mother of modern health care quality. A contemporary of Semmelweis, she introduced sanitation and hand washing at the bedside while serving at the Barrack Hospital in Scutari, Turkey during the Crimean War. This simple intervention decreased mortality from 50% to 2%. She returned to England and with William Farr, a statistician, published the outcomes of London hospitals. They noticed that hospital deaths were dependent on the patient’s age and condition at the time of admission to the hospital. Nightingale’s quality improvement legacies were the emphasis on
sanitation at the bedside and the importance of risk adjustment when assessing outcome.\(^5\)

Earnest Amory Codman is, arguably, the father of health care quality. His major contribution was a careful tracking of the outcomes of surgical procedures at the Massachusetts General Hospital. If an adverse outcome occurred, Codman established a methodology for investigating the cause and tracking it as due to either patient disease or due to an error in management.\(^6\)

Progress in advancing quality improvement in health care languished through the Great Depression and World War II. After the War, medicine changed rapidly, eventually becoming primarily hospital based. With that came hospital performance assessment and accountability, emphasizing care based on evidence.\(^7\) At about this time, William Deming, a PhD from Yale in electrical engineering and statistics, was sent by our government to Japan to resurrect their industrial complex, which had been destroyed by incendiary bombing and two nuclear weapons. He applied the Shewart cycle to the Japanese industrial process—essentially a continuous repeating cycle of plan, do, check act, and evaluate, which is now referred to as the “Deming Cycle.”\(^8\) Examined closely, this is remarkably similar to the scientific method attributed to Francis Bacon in 1620: hypothesize, experiment, and evaluate. Deming’s success ultimately led to the application of the Shewart cycle to hospital quality programs. Arguably, one of his major contributions to improving quality in industry was eliminating handoffs.\(^9\)

The last major contribution to improving hospital performance is the analysis of “value” as advocated by Michael Porter, a PhD in business economics from Harvard. Simply put, value equals the quality of the process plus the quality of the outcome divided by the cost.\(^10\) Porter’s work completes the necessary “primer” on quality in health care to support my thesis. Now, let’s move on to evidence based medicine.

William Blake, poet and author, wrote: “Reason, or the ratio of all we have already known, is not the same that it shall be when we know more.” Similarly, from Francis Bacon, “So let great authors have their due, as time, which is the author of authors, be not deprived of his due, which is further and further to discover the truth.” These quotes emphasize the importance of substantiating, beyond reasonable doubt, all theories (as a resident, I was doing major gastric surgery for peptic ulcer disease, almost every week for, what turned out to be, an infectious disease. The truth, which we now know, but did know then, is the vector H. pylori.) My favorite quote comes from Rudyard Kipling: “I keep six honest serving men (They taught me all I ever knew.) Their names are: what and where and when and how and why and who.” Kipling realized the power of human emotion in developing a bias when he added the “who.” Unfortunately, we can delude ourselves if we have a strong bias.

Evidence-based medicine is intended to optimize decision making by emphasizing the use of evidence from well-designed and well-conducted research, the very basis of which is the scientific method. Theory, on the other hand, is speculation, or conjecture, that is unsubstantiated. It may be reasoned, based on empiricism or logic, but it is still unproven without an evidence base. Based on current knowledge, it may even lack face validity.

Let us now discuss the first theory: “Limitation of resident work hours will improve the quality of health care at academic institutions.” The event that incited this theory was the death of Libby Zion, an 18-year-old college student who presented to a New York City hospital in 1984 with fever and some shaking chills.\(^11\) The admitting diagnosis by the house officers, in concert with Libby’s private physician, was “viral syndrome with hysterical symptoms.” Over the next 7 hours she rapidly deteriorated, suffered a respiratory arrest and could not be resuscitated. Based on her clinical course, her medication history, the drugs that she received after admission, and the postmortem examination, she was judged as having serotonin syndrome, a symptom complex associated with the use and interaction of neurotransmitter uptake inhibitors. At the time of Libby’s death, little was known about the syndrome.\(^12\)

That summarizes the “what,” the “where,” the “how,” but not the “who.” The “who” is Sidney Zion, Libby’s father. Zion had been a trial lawyer and was past Assistant US Attorney for New Jersey. He was also a journalist for the New York Times. After being fired from the Times, he revealed the confidential source of the Pentagon Papers, which were then being published by the Times and the Washington Post.\(^13\) For what appeared to be a vindictive ethics breach directed at the Times, he was considered by many to be a “pariah” among journalists. Bereaving after his daughter’s death, Zion wrote an op-ed piece in the Times implicating the long hours of resident work as a factor affecting her care and ultimate outcome. Zion pressed the District Attorney to convene a grand jury and indict the house officers for murder. A grand jury was convened and refused to indict the house officers for murder, but did charge them with 38 counts of gross negligence in Libby’s death. Under New York law the investigative body for such charges is not the grand jury, but the Hearing Committee of State Board of Professional Medical Conduct. The committee agreed to hear the case. Over a period of 17 months and 30 hearings, 33 witnesses testified, including experts in toxicology, emergency medicine, and internal medicine. Among the witnesses were chairmen of internal medicine from six prominent medical schools, some of whom testified under oath that they had never heard of the interaction between the drugs administered in the hospital causing serotonin syndrome. At its conclusion, the committee unanimously decided that the evidence presented supported none of the 38 charges against the house officers. Meanwhile, the New York State Health Commissioner decided to investigate “systemic problems” in residency training and convened a blue-ribbon panel chaired by Bertrand Bell, a primary care physician and advocate for resident supervision. Before the conclusion of the Hearing Committee deliberations, the Bell Commission made several recommendations: increased emergency room staffing, supervision of in-house residents by an attending physician, legislation regarding the use of restraints, and a computerized system to warn of drug interactions. All of these were reasonable and based on the facts of the case. However, without any evidence to suggest that fatigue or the duration of shifts worked by the residents were a factor in causing Libby’s death, the commission promulgated regulations to limit work hours for interns and junior residents: 12-hour shifts for those who worked in the emergency department and 16 hours for those who worked on the hospital floors. For the 50 New York City hospitals, this required the hiring of 2,045 physicians and 974 ancillary staff, at a cost of almost US $204 million. If these promulgations were supposed to represent a quality initiative, one would expect that
the blue-ribbon panel would have identified some outcome measures (i.e., plan, do, CHECK, act), but there were none specified. Arguably, the implementation of shift work and "handoffs" could, according to Deming, decrease quality. Also, the cost, without specified measures of an expected outcome, would diminish value, according to Porter.

The Bell Commission's recommendations had a penumbra that extended far beyond New York State. In 1990 the Accreditation Council for Graduate Medical Education (ACGME) established an 80-hour workweek for medical specialties. In 2001, resident organizations petitioned the Occupational Safety and Health Administration to limit work hours for residents and fellows as a "workplace safety issue" (not a patient safety issue). Congress threatened legislation, and in 2003, the ACGME issued an 80-hour workweek requirement for all specialties. This was done without an evidence base suggesting that patient or resident workplace safety would be impacted by these recommendations. It remained a theory, untested conjecture.

That the theory might be wrong became apparent almost immediately. As early as 2004, resident handoffs were considered a "perilous exchanges" leading to adverse events and this "handoff problem" continues. To determine the current evidence base for the efficacy of resident work hour restrictions (RWHR), I did a nonsystematic review of the literature. The primary outcome of interest was quality as measured by all cause mortality. I limited the search to the United States and to medical and general surgical care beginning in 2005. I reviewed 122 abstracts and did a full text review of 28 articles. For brevity, I will discuss four, which I consider to be the most salient.

The first was the report of the Institute of Medicine (IOM) published in 2009. The authors of the IOM report considered findings regarding patient outcomes to be inconclusive due to the "poor quality of the existing literature and the lack of reliable data to track outcomes." The authors were "disappointed with the lack of any comprehensive attempt to detect changes in residency programs and their effect on education and patient safety." The available data suggested, "slight, but inconsistent, decreases in mortality of medical, but not surgical, patients." The estimated added cost of provider substitution was US $1.7 billion for a programmatic change without reliable evidence.

To address the poor quality of the literature identified by the IOM, Philbert and colleagues published a review in 2013 selecting only those studies of "high quality," as determined by the Medical Education Research Quality Index. Thirty-four studies of 1,515 reviewed qualified for inclusion based on a Medical Education Research Quality Index score greater than 12. Only nine (26%) of them demonstrated improvement while 25 (74%) reported no change or worse outcomes. Only 2 (10%) of 20 studies reporting on surgical outcomes reported improvement. Morrison et al. demonstrated a reduced mortality from 5.16% before implementation of RWHRs to 5.03% after ($p < 0.03$) on data from over 500,000 patients treated at academic centers reporting to the National Trauma Data Bank. The only other "high quality" work reporting on surgical patients came from Privette and co-workers. This single-center study of over 14,000 patients with data collected concurrent with care demonstrated a significant reduction in mortality from 1.9% to 1.1% ($p = 0.002$) and a reduction in provider related complications ($p < 0.001$) after implementation of RWHR. Of note, the RVU-82 billing modifier (No Qualified Resident Available) increased more than 10-fold, from 523 pre-RWHR to 6542 post-RWHR. This suggests that the improvement was not necessarily due to a reduction in resident hours, but to an increase in the amount of work done by attending surgeons, which was confirmed by review of clinical hours reported on attending surgeon Medicare audit time allocation. There was additional cost in terms of salary and fringe for new hires, resulting in a decrease in value.

Based on this nonsystematic review, the evidence to support improvement in patient safety following RWHR is weak at best and certainly not uniform among specialties. The apologists for RWHR may suggest that resident education is improved and lifestyle is better. The lifestyle issue is moot; if you reduce my hours, but do not reduce my pay, my lifestyle is better. On the issue of surgical resident education, two recent studies are germane. Bilimoria and colleagues compared a flexible hour motif (waivers for time between shifts and maximum shift length, but adherent to 80 hours averaged over four weeks) to the standard ACGME requirement while examining patient safety and a variety of outcomes including resident satisfaction with education. There were no significant differences in patient safety outcomes and resident satisfaction with education. Interestingly, residents in the flexible motif were less likely than their counterparts on the standard ACGME schedule to perceive a negative effect of RWHR on patient safety, continuity of care and professionalism. Finally, Mattar et al. reported on a survey sent to fellowship directors inquiring about the preparedness of recent graduates of general surgery residency programs for fellowship training. The respondents listed major deficits in patient ownership, psychomotor ability, ability to operate independently, and ability to manage postoperative problems. Dr. Frank Lewis, then Executive Director of the American Board of Surgery, commenting on the presentation at the American Surgical Association made very pertinent points that are worth quoting in full, "I have no substantial argument with the points being made, but would note that the issue of inadequate preparation should not be considered newsworthy. The American Board of Surgery has held retreats and has addressed resident preparation on multiple occasions over at least the past decade. On the board certification examination, the failure rate on the oral examination has been progressively rising, from 16% in 2006 to 28% in 2012. There has been a relatively consistent failure rate, in the high teens, on the written examination. The failure rates on the two examinations do not substantially overlap, so that the results are more or less additive. Therefore, the total number of people who fail either the written or the oral examination for the first time is typically around the middle thirties. This rate is arguably far too high for a group of talented people who should have mastered the subject after completing a five-year training program."

I believe that RWHR in surgery was an emotional reaction, not a thoughtful response, to a constellation of issues, including a declining medical student interest in surgery. As a quality initiative, the process was severely flawed; there was no plan, no pilot project (do) no means of assessing outcome (check), and no mid-course correction (act). As a result, it did not impact patient safety either way. Furthermore, surgical
residents lost the opportunity to follow a disease process from beginning to end (due to handoffs in care when a shift ends) resulting in not only less confidence in their abilities, but also poorer performance on standardized tests.

Now, the second theory: “Pain is the fifth vital sign.” Pain, as we all know, is not a sign, it is a symptom. Signs (i.e., blood pressure, pulse, etc.) are objective and measureable; symptoms are subjective, difficult to assess, and impossible to quantify or measure objectively. As a symptom, pain is similar to hunger. If you ask me right now, Steve Shackford, on a scale of one to ten, how hungry are you, I would say, well, now that you mention it, it’s a four, and I think I would like a doughnut!

In retrospect, it seems ludicrous that such an absurd concept gained momentum. How did this happen? Baker,24 writing as a member of the Division of Healthcare Quality Evaluation of the Joint Commission for the Accreditation of Healthcare Organizations, provided an account. Baker cites strident calls for better pain management beginning in 1973, based not only on patient's reports of inadequate pain relief, but also on the reported low potential for addition when opiates were used. In reviewing Baker's references (and the references of his references), I found a significant lack of rigorous study design and follow-up and would classify the majority of the citations I reviewed as Level 5 based on the limited control of confounding, heterogeneous populations, more than minimal bias, and, in most cases, a lack of follow-up to assess addiction.25 For example, Baker credits Mitchell Max's commentary26 with calling attention to the need to improve the treatment of pain in hospitalized patients. In support of his contention of untreated or undertreated pain, Max cites the “classic study” authored by Marks and Sachar.26 This “classic study” consisted of a single 15-minute interview and a chart review (to confirm diagnoses and record analgesic drug doses) of 37 hospitalized patients. The interviews were performed by multiple members of the psychiatric “liaison staff” using an validated pain assessment scale. Severe pain persisted in 12 (37%) and moderate pain persisted in 15 (41%). An assessment of interrater reliability was not done, no information is given on the time between the last dose and the interview, and there was no follow up to see if the pain assessment changed with “adequate” medication. Another study evaluated pain assessment in 156 patients in the emergency department at a Level I urban trauma center.27 The authors noted a significant disparity ($p < 0.001$) between the nurses' assessment of pain and the patients' assessment. Based on the standard deviations of the pain assessment scores, there was overlap indicating, as the authors point out, that there was agreement in a number of cases. There was no validation of the patient's cause of pain (i.e., by diagnostic studies or imaging), no assessment of pain relief after the nurses provided medication, and there was no follow-up on return visits. However, the most glaring example of the inadequacy of the literature is a letter to the editor of the New England Journal of Medicine describing pain management in 39,946 hospitalized patients “monitored consecutively” (not described in any detail) with “only four cases of reasonably well documented addiction.” The authors concluded, “that despite widespread use of narcotic drugs in hospitals, the development of addiction is rare in patients with no history of addiction.”28

This letter was cited 628 times (often uncritically) and helped “to shape the narrative that allayed prescribers' concerns about the risk of addiction associated with long-term opioid therapy.” There was a spike in citations following the introduction of OxyContin in 1995.29 Of note, in 2007 the manufacturer of OxyContin and three senior executive pleaded guilty to criminal charges that they misled doctors, patients and regulators regarding the drug's addictive potential.30 In 2000, based on the available literature and recommendations of the American Pain Society, JCAHO proposed standards for pain management using a validated numeric scale. Soon, however, concerns were raised about overzealous use of opiates as the number of “over-sedation episodes” increased.31 JCAHO began a series of de-escalation recommendations from “pain is the fifth vital sign” in 2001 to “pain used to be considered ...” in 2002 to complete elimination of the phrase from the accreditation manual in 2003.

What were the unintended consequences? Since 2000, there have been 183,000 deaths due to drug overdose.32 This is now recognized as a health care crisis demanding intervention by local, state, and federal governments.

I will close with another, but more current, initiative from New York State: “Rory’s Regulation,” which has undeniable similarities to the “Libby Zion Law.” Rory Stauton was a 12-year-old boy who died of sepsis following a soft tissue infection. Rory's Regulation is a state mandate that all hospitals use evidence-based protocols for sepsis identification and treatment. Though the mandate is still in its “infancy” the state issued a report that sepsis protocol use increased and the mortality from sepsis decreased leading Governor Cuomo to issue a press release that the state's efforts were “saving lives.”33 Sepsis-related mortality is decreasing across the nation and without data from New York State prior to the issuance of the regulation it is impossible to say if it has had any significant impact or simply mirrors the decline seen in other states during the same time period. Similar to the transference of bereavement into action exhibited by Sidney Zion, the Stauton family has formed a foundation that is actively seeking sepsis-protocol mandates in every state by 2020.

We, all specialties of medicine, must be at the table as new regulations or mandates come forward that will affect our autonomy in caring for our patients. We must assure ourselves and our patients that any mandated changes in protocols are responsive to an evolving evidence base and that these mandates improve the quality and the value of health care. Our patients expect that of us as part of the social contract.

DISCLOSURE
I have nothing to disclose. This work is my own and I received no services or remuneration for writing the manuscript for publication.

REFERENCES


