Editorial

Reflections on Improving Hospital Performance

The juxtaposition of preparing an editorial for the October 2008 volume of *Health Services Research (HSR)* based on several studies about performance measures in healthcare organizations and participating in the hectic last weeks of Marie-Claire Rosenberg's thesis (ABF is her thesis supervisor) on nursing care and hospital performance (Rosenberg 2008) triggered a revelation. By coincidence, it was exactly 30 years ago that Ann Barry Flood published the first article from her thesis on hospital structure and performance (Flood and Scott 1978) and was a coauthor on the last major report on quality of care in US hospitals from the Stanford Center of Health Care Research (SCHSR) (Forrest et al. 1978).

What has changed since the SCHSR's reports and how is the field advanced by current work in this issue?

Thirty years ago, in follow-up to a provocative article that suggested that American hospitals varied dramatically in their quality of surgical care, SCHSR was among the first to use detailed risk adjustments based on patient level overall health and treatment to "adjust" for outcomes in order to assess hospital-level performance. The basic design was an overlapping study with relatively scant information available about a large sample of 1,224 hospitals (using the American Hospital Association's [AHA] annual survey) and some 600,000 of their surgical patients (using computerized abstracts from medical charts) in contrast to a detailed prospective study of 10,000 surgical patients followed in 17 of these hospitals where SCHSR collected seven forms about each patient's outcomes and postoperative treatment and interviewed about 80 people and surveyed several hundred physicians and nurses at these hospitals.

The similarities to today's work? The Center's work focused on outcomes as the "ultimate" indicator of quality, in part because there was not much evidence that process measures led to substantial improvements in health—only for a handful of measures. A chief concern about using outcomes was how to reliably measure them and adjust for risk factors, because those outcomes that were most reliably measured (death) tended to also be rare. Organizational factors studies tended to focus on the importance of

physicians, both as the surgeons in charge of individual patients and as medical staff with responsibility for peer review and professional oversight. Not surprisingly, we too found important variation in quality in hospitals—variation in outcomes that remained after stringent attempts to adjust for patient risk factors in both the large-crude study and small-intensive study.

Two “surprises” we found were in regard to organizational factors related to better quality: (1) The more explicit were the policies and procedures for the nursing staff, the better were the outcomes. (This finding seemed to contradict the professional model that implied that flexibility and judgment should be left to individuals; instead, we found that rigid rules and regulations—such as those that promote safety and prevent equipment failure or loss—were important.) (2) Hospital-based factors explained more of the variation in adjusted outcomes than did surgeon experience and training. (This finding was examined very carefully by the physicians on the team but held in test after test—including the association between larger volume of cases of a particular type of surgery and better outcomes.)

What did we “forget” to look for 30 years ago? No one looked at the impact of *system membership* or *vertical integration*—these were not even questions on the AHA survey at that time nor in organization theory about non-profit service sectors; we also paid no attention to *health maintenance organizations* (there were not many and almost all charges were on a fee-for-service basis); *market penetration* or *environmental influences* (one hospital administrator noted that the only reason he paid attention to other hospitals is that otherwise he had no idea what to charge); and *quality improvement*—newly re-introduced to businesses in the United States—was thought not to apply to healthcare.

Other differences: In order to run our logistic regressions to adjust for patient risk factors, we shipped our programming code across the country to one of the biggest mainframe computers, which took almost 48 hours to converge on five regressions involving 600,000 patients. We were also limited to the approximately 1/5 of the nation’s hospitals that participated in an electronic system of patient chart abstracts. Neither claims nor medical records nor most inventory records were computerized, so that a practical way of monitoring hospitals or paying for better performance was unimaginable.

Turning to the studies in this issue, five are of particular interest to this story. They continue the quest to measure quality and hold hospitals accountable for it; they look at measures reflecting nursing care in hospitals; and they look for practical ways well-intentioned managers and providers can improve the quality their organizations provide.

THE QUEST TO MEASURE QUALITY

A general aphorism of proponents of improving health care is that we need to start with *measuring* quality in order to improve it, to design incentives to reward it and to inform consumers so that they can choose better healthcare organizations and providers. While outcomes have long been regarded as the “ultimate” measure of quality, various complexities in measuring them (including usually focusing on relatively rare “failures” rather than successful outcomes and their sensitivity to patient risk factors and general health) means that most pay-for-performance and consumer report cards focus on process measures instead. However, the ultimate goal is still to improve the health outcomes—so is a concentration on process measures justified?

Rachel Werner, Eric Bradlow, and David Asch set out to address this question (Werner, Bradlow, and Asch 2008). They posit that measuring how well hospitals do in performing evidence-based processes of care will improve care in two ways. First, patients whose care is directly linked to the measured processes will benefit. Second, they argue that hospitals that perform particularly well on these measures will also be set up to provide better care in general and so these measures can serve as a more general marker of quality. They tested these ideas by estimating how much a hospital’s death rate would be improved by how well it performed on three evidence-based process measures. Their findings suggested that performance measures were indeed identifying hospitals whose performance—as indicated by patient outcomes—was better *overall* and not merely better for the patients having conditions with *measured* quality markers.

While Werner et al.’s work suggests that process measures may provide an important “edge” over the more complex outcome measures for rewarding performance—at least at the hospital level, Alan West, Bill Weeks, and James Bagian’s article suggests a cautionary note (West, Weeks, and Bagian 2008). They focus their efforts on the patient safety measures that are basically intermediate outcomes, potentially preventable with good nursing and physician care (e.g., failure to thrive, falls during hospitalization, decubitus ulcers that arise during hospitalization, and postoperative infections and other postoperative complications). These safety “outcomes” are arguably the outcomes of poor process of care during hospitalization and not patient outcomes that reflect “risk factors” that arise from admitting sicker patients. This article suggests that, because these factors tend to be rare events, they too are relatively unstable in assessing hospital performance. They found that, while regional-level rates might suggest the need to investigate quality/process

measures, the hospital-level rates were too unstable to assess “poor performance” measures that deserve performance-reprimand.

MEASURING NURSING CARE IN HOSPITALS

Joanne Spetz, Nancy Donaldson, Carolyn Aydin, and Diane Brown took a different tact to examine research evidence and performance-based measures of nursing “quality” (Spetz et al. 2008). Nursing workforce measures, including percent of Registered Nurses in hospital-based care, have long been used as hospital based measures of “structural” quality—i.e., the hospital’s capacity to deliver good quality care. The authors took advantage of four data sets available to describe nursing staff in California hospitals to compare and contrast several typical measures of nursing workload.

Remarking on reports of contradictory “facts” reported in the literature as well as discrepant research findings, they explain confusion arises in part due to differences in recording and reporting nursing workload. Here are five examples:

- Hospital payroll and nursing directors on units can differ on what hours they report as *direct* nursing care, the first focusing on avoiding sick or vacation hours while the other focuses on avoiding administrative duties.
- Measures can lump registered nurses, licensed practical nurses and nursing assistants together or may report them separately or only combine the first two.
- Hospitals may be asked (such as by the AHA) to report nurses in all associated units, including any working in long-term care or out-patient units.
- Measures that combine part- and full-time nurses to estimate fulltime equivalent effort often make assumptions that mis-estimate the true number of hours worked per week (both for full- and part-time nurses) and may overlook float nurses.
- Measures that report workload in terms of patient days can also misrepresent true workload for two reasons: (1) a “patient day”, as recorded for purposes of reporting a patient’s total length of stay, often reflects stays that may be considerably shorter *or* longer than 24 hours; and (2) units that apparently have the same number of beds filled on a given day can arguably have different “workloads”

depending on the intensity of care needed, including whether there was a high turnover of “new” patients filling those beds.

The authors then examined typical nursing workforce measures in four databases with different ways of recording nurses and hours worked. Despite the differences in data reporting, they found strong consistency in portraying hospitals. The greatest discrepancy was between staffing as reported at the unit level and then aggregated or reported directly at the hospital level. They conclude that no reporting method was “better” than all others, but caution researchers to be more careful so that it is clear what data were included and with what implications for their findings.

EVIDENCE ABOUT STRATEGIES FOR IMPROVING QUALITY

These last two articles illustrate how difficult it is for healthcare organization to implement evidence-based changes to improve the quality of their care. Both sets of researchers attempted to evaluate “real-world” settings where the leadership and resources and actions emanated from the organizations themselves rather than being undergirded or provided totally by the researchers.

Elizabeth Yano, Lisa Rubenstein, Melissa Farmer, Bruce Chernof, Brian Mittman, Andrew Lanto, Barbara Simon, Martin Lee, and Scott Sherman evaluated a quality-improvement based effort to help Veterans quit smoking (Yano et al. 2008). Eighteen VA facilities were randomized to receive quality improvement assistance (structured evidence review, local priority setting, quality improvement plan development, practice facilitation, expert feedback and monitoring) or mailed guidelines and VA audit-feedback reports. Yano and her colleagues then evaluated the success of these practices to get Veterans to quit smoking and what changes in practices occurred. While intervention practices saw an increase in the number of referrals (which was not recommended by the experts or evidence review), there was no effect on quit rates. They also found that the “intervention facilities” had intended to implement several complex and recommended changes, but most attempts were never successfully implemented except for increasing referrals. What was particularly interesting, and pertinent to our discussion, was the authors’ reflection on these results: The practices, despite expert advice and evidence reviews to the contrary, implemented referrals in part because they still believed in their value and in part because it required the fewest resources and

least effort. Without new resources or reward for additional work, improvement and change is very hard to accomplish.

In a different approach, Janneke Grutters, Manuela Joore, Frans van der Horst, Robert Stokroos, and Lucien Anteunis provide an example of how decision makers in organizations, using their proposed framework, can evaluate whether a major change in how their organization delivers care would be cost-effective to implement (Grutters et al. 2008). Echoing an observation made in *HSR* by Jack Hadley (2000), the authors say that there is rarely good enough data to allow decision makers to make timely and precise enough calculations to evaluate whether an innovation is worthwhile. Moreover, most organizations are not convinced enough about the “cost-effectiveness” of cost-effective analyses to devote the resources needed to gather data.

Grutters et al. proposed a Markov chain variant of decision analyses that, in their example, allowed decision makers in the organization to fairly efficiently examine the cost effectiveness of the change and to explicitly examine whether there were any apparent untoward decreases in patient safety.

As we can see from the examples above, the field has developed, evolved, and changed. Of note is the change to use performance measures to hold organizations accountable for their quality and to try to use evidence-based approaches to improve care. Perhaps, the one thing that has remained consistent: we still need to improve our measures of performance and our advice to organizations about how to improve their care.

Ann Barry Flood and Marie-Claire Rosenberg

REFERENCES

- Flood, A. B., and W. R. Scott. 1978. “Professional Power and Professional Effectiveness: The Power of the Surgical Staff and the Quality of Surgical Care in Hospitals.” *Journal of Health and Social Behavior* 19 (3): 240–54.
- Forrest, W. H., B. W. Brown, W. R. Scott, W. Ewy, and A. B. Flood. 1978. “Impact of Hospital Characteristics on Surgical Outcomes and Length of Stay.” Final Report. National Center for Health Services Research, Department of Health, Education, & Welfare (DHEW). pp. 438.
- Grutters, J. P. C., M. A. Joore, Horst F. van der, R. J. Stokroos, and L. J. C. Anteunis. 2008. “Decision-Analytic Modeling to Assist Decision Making in Organizational Innovation: The Case of Shared Care in Hearing Aid Provision.” *Health Services Research* 43 (6): 1662–73.
- Hadley, J. 2000. “Better Health Care Decisions: Fulfilling the Promise of Health Services Research.” *Health Services Research* 35 (1 Pt 2): 175–86.

- Rosenberg, M. -C. 2008. "Ph.D. Thesis: Magnet Hospitals and Patient Outcomes." The Dartmouth Institute for Health Policy and Clinical Practice. Hanover, NH: Dartmouth College.
- Spetz, J., N. Donaldson, C. Aydin, and D. S. Brown. 2008. "How Many Nurses per Patient? Measurements of Nurse Staffing in Health Services Research." *Health Services Research* 43 (6): 1674–92.
- Werner, R. M., E. T. Bradlow, and D. A. Asch. 2008. "Does Hospital Performance on Process Measures Directly Measure High Quality Care or Is It a Marker of Unmeasured Care?" *Health Services Research* 43 (6): 1464–84.
- West, A. N., W. B. Weeks, and J. P. Bagian. 2008. "Rare Adverse Medical Events in VA Inpatient Care: Reliability Limits to Using Patient Safety Indicators as Performance Measures." *Health Services Research* 43 (6): 1737–51.
- Yano, E. M., L. V. Rubenstein, M. M. Farmer, B. A. Chernof, B. S. Mittman, A. B. Lanto, B. F. Simon, M. L. Lee, and S. E. Sherman. 2008. "Targeting Primary Care Referrals to Smoking Cessation Clinics Does Not Improve Quit Rates: Implementing Evidence-Based Interventions into Practice." *Health Services Research* 43 (6): 1637–61.