

**Testimony to**

**The United States Senate Committee on Finance  
Subcommittee on Health**

**A hearing addressing:**

**What Is Health Care Quality, and Who Decides?**

**Wednesday, 18 March 2009, 2:30p  
Dirksen Senate Office Building, Room 215**

**Brent C. James, M.D., M.Stat.  
Member, National Academy of Science's Institute of Medicine  
Chief Quality Officer *and*  
Executive Director, Institute for Health Care Delivery Research  
Intermountain Healthcare  
36 S. State Street, 16<sup>th</sup> Floor  
Salt Lake City, Utah 84111-1633  
(801) 442-3730  
Brent.James@imail.org**

## **Quality Measurement and Improvement: A Tale of Two Projects**

### Project 1

In December, 2004, the Institute for Healthcare Improvement launched its 100,000 Lives Campaign. The Campaign targeted 6 clinical areas: central venous IV line infections, ventilator-associated pneumonias, rapid response teams, surgical site infections, acute myocardial infarction (heart attack) care, and medication reconciliation. For each target, IHI developed “bundles” that described implementable best care processes at the bedside. Each bundle included embedded process and outcome measures. IHI sponsored teleconferences, web sites, and meetings in which hospitals shared barriers, lessons learned, and successes. More than 3,100 U.S. hospitals, representing more than 80 percent of all U.S. hospital admissions, voluntarily participated. They were motivated by professional values – better patient care – without direct financial incentives or regulatory mandates.

IHI estimated that the 100K Lives Campaign saved over 120,000 lives. While the overall IHI evaluation method was somewhat controversial – the IHI Campaign approach lacked the structure necessary for careful observational research – a large number of individual hospital-level instances of improvement from within the Campaign are compelling. For example, a Johns Hopkins University Hospital team helped the Michigan Hospital Association implement the central venous line infection bundle in 108 ICUs (Pronovost et al., NEJM, 2006). Historically, about 80,000 such infections occurred in U.S. hospitals each year, accounting for more than 28,000 deaths and increasing health care costs (marginal resource consumption to treat the infections) by more than \$2.3 billion. The Hopkins team supported the local clinicians with embedded measurement and implementation advice. Central line infection rates fell from 7.7 to 1.4 per thousand catheter days. Mortality rates fell by more than 1,700 deaths per year in Michigan alone.

### Project 2

In 2004, the DHHS Centers for Medicare and Medicaid Services (CMS), acting under Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA), reduced Medicare payments to hospitals for certain diagnoses by 0.8 percent. Hospitals could recover the payment by reporting 10 “quality performance measures.” Over time, CMS has added additional measures. The core set now addresses 5 areas: Acute myocardial infarction (heart attack) (7 measures); heart failure (4 measures); community-acquired pneumonia (7 measures); infection, hyperglycemia, and venous thromboembolism prophylaxis in surgery (7 measures); and asthma care in children (2 measures).

Subsequent external evaluations showed that while hospitals’ performance scores on the CMS process measures changed significantly, final outcomes did not (Werner & Bradlow, JAMA, 2006; Fonarow et al., JAMA, 2007; Managlat, Smith, & Butler, J Card Fail, 2008).

Why the difference? How is it that a privately launched, voluntary effort produced massive improvements in quality, while a parallel governmental effort produced, at best, a very minor impact?

### Background information

#### **Quality measurement has improved significantly over the past 3 decades:**

- W. Edwards Deming linked quality to underlying work processes. He suggested that every process produces 3 parallel classes of outcomes: quality, cost, and service. This provided a robust structure for quality measurement, in context.
  
- Health services researchers (Nelson, James) further broke medical quality into 4 major subdivisions, which greatly simplified measurement within much more consistent categories. Those 4 major subdivisions are:
  1. appropriateness (indications)
  2. complications
  3. therapeutic goals (biologic performance as seen by a health professional)
  4. patient functional status (biologic performance as seen by a patient)
  
- These advances have led to validated quality measures within well-defined patient populations.

#### **Despite those advances, quality measurement still has major limitations:**

- There are widespread problems with incomplete science, incomplete assessment, incomplete documentation, and incomplete data extraction from fragmented, dispersed medical records.
- “Availability bias”
- Problems with attribution (most care is delivered by teams)

#### **Any quality measurement system itself contains variability, which can obscure the underlying care delivery performance:**

- there is a clear need for feedback and follow up on the data system itself, using well-established methods found in industrial quality control theory (gauge theory)
- no national groups currently employ this critical element
- example of how it works: condition-specific measurement within Intermountain Healthcare

#### **As a result, it is currently impossible for quality measures to accurately rank providers in most circumstances:**

- a very robust scientific literature supports this conclusion (will supply on request)
- good quality accountability therefore needs to use approaches that do not rely on ranking – these approaches do exist, primarily derived from quality improvement theory

**Provider quality performance is highly condition specific:**

- 3 decades of investigation have found no reliable general quality indicators (the fact that a provider does well or poorly on one condition does not imply that the same provider will do well or poorly on other conditions)
- however, care delivery concentrates massively. About 10% of clinical conditions account for over 90% of all care delivery
- therefore, build in measures by condition, in size order, to address the most good for the most patients

**Poorly-constructed quality measurement systems often lead to “data gaming”**

(principle: it is easier to look good than to be good):

- There are 3 ways to get a better number (Deming):
  1. improve the underlying process
  2. shift resources to the area under the measurement spotlight, at the expense of areas not under the measurement spotlight (very often, the peripheral damage outweighs the focused gain)
  3. game the number
- “as one attaches greater rewards or punishments to achieving a number, one gets increasing proportions of (2) and (3)”
- extrinsic rewards tend to destroy intrinsic motivation
- it is very clear that type (2) and (3) activities are becoming common among U.S. hospitals, relative to the CMS measures

**Transparency is not the same as accountability:**

- high-quality care delivery usually involves a series of decisions around sequential care delivery choices
- patients usually make those decisions in the context of a caring relationship, with a physician or nurse advisor
- “transparency” means that all participants – the clinician advisors as well as the patients – have sufficiently accurate, detailed information to make wise choices at each step in the chain
- Accountability measures, that reduce the problem to a single patient choice of a hospital or a physician, can directly undermine the true transparency that is essential to high quality care.

**There are 2 primary approaches to quality - (1) measurement for selection (accountability) versus (2) measurement for improvement:**

- measurement for improvement contains measurement for selection / accountability – the opposite is not true (measures for accountability, mandated from above, do not create capacity for actual quality management and improvement at the front line)
- measurement systems designed for accountability often consume limited front-line resources and actively damage quality of care (Localio, NEJM, 1999; Wachter et al. Ann Int Med, 2008)
- there are rigorous methodologies for generating reliable front-line, embedded data systems that minimize burden and maximize data quality (NQF SFB report). These

methods stand in contrast to the political methods currently used by most national reporting groups.

**Two national groups are showing strong approaches to quality measurement:**

- NCQA methodology has historically been strong
- The new Joint Commission initiative holds very great promise

**The CMS and Joint Commission ORYX measures currently face significant technical challenges:**

The CMS measures operate in parallel with other quality measures required by The Joint Commission (the ORYX system). A hospital can submit its measures directly to CMS through a web-based interface (the CART system). However, The Joint Commission requires data submission through tested and certified Performance Measurement System Vendors (Vendors). As most Vendors offer parallel support for both CMS and The Joint Commission ORYX system, most hospitals combine the two activities into a single effort, using the services of a certified vendor. Once CMS has received a hospital's quality measures, either through direct submission (CART) or a Vendor, CMS performs computerized integrity checks on the submitted data.

Over time, the CMS measures have become operationally complex. For example, Heart Failure is arguably the simplest of the CMS quality areas. It includes 4 main measures, along with patient demographic data. Evaluating the 4 main measures requires evaluation of almost 20 data subelements. Each of the subelements has complex descriptions – inclusion and exclusion criteria – that run from 1 to 6 pages per in length.

Both CMS and The Joint Commission have used the National Quality Forum (NQF) to select appropriate measures for quality reporting. NQF has established committee structures that represent major constituencies, including health insurance groups, patient advocates, system vendors, health and hospital professional organizations, and care providers. The measure selection process usually starts with a review of available medical evidence, then uses a political consensus approach that draws upon the various constituencies.

When constructing its quality measures, CMS faced a major dilemma: While “quality” innately implies final outcomes that are important to patients – for example, mortality following a heart attack – it is very difficult to account for underlying differences in patients (severity of illness or risk adjustment) when interpreting final outcomes. CMS therefore chose instead to track process measures (also known as intermediate outcomes) – the factors that current best medical science suggest should drive final outcomes. This had the effect of greatly increasing effective sample sizes, and of shortening measurement timelines. CMS relied upon process measures that showed a strong association with final outcomes in the medical literature (evidence-based medicine). The key to using process measures is the strength of the linkages between the intermediate outcomes used for performance assessment, and the final patient outcomes that the intermediate measures are believed to predict. Unfortunately, those linkages are highly sensitive to small changes in the underlying data systems. As the complexity of the

underlying data definitions increases, opportunities to change measured performance by purposefully or inadvertently manipulating the underlying data system multiply.

As result of the foregoing, both the CMS measures and The Joint Commission measures are designed for “after the fact” chart abstraction (as opposed to being embedded into the clinical work flow); they rely on a subset of measures based upon what can be found in a typical hospital chart or existing electronic financial data systems (called “availability bias”), and focus on process measures rather than final outcomes measures. Subsequent external evaluation suggests that the process-outcomes linkages upon which the CMS measures rely are weak.

### Recommendations

1. **Build balanced measurement** (clinical intermediate and final outcomes, cost outcomes, and service outcomes) **for specific clinical conditions, in priority order**. Prioritize on the basis of careful analysis addressing (1) the number of patients affected, (2) risk to the patient (= intensity of care = cost per case), (3) internal variability (coefficient of variation in care intensity, within a condition), and (4) social equity (underserved populations).
2. **Build the measures from the bottom up** – create a measurement set that can embed in care delivery at the bedside, and that directly supports the ability of clinical teams to deliver care, manage care processes, and systematically improve. **“Roll up” selected front-line measures into system, State, and national reporting.**
  - this approach minimizes burden on front-line teams
  - it is the best way to insure accurate, complete, and timely data (by centering around data that are actually used at the point of patient contact)
  - it provides true transparency – it will inform all involved in the chain of clinical decision making (patients, physicians, nurses, etc.)
3. **Examine outlier cases to find root causes, then use the resulting knowledge to systematically clean up and improve the data system itself** (gauge theory). For example:
  - Intermountain Healthcare currently has a full set of intermediate and final, clinical, cost, and service outcomes measures available to patients, clinicians, and managers, for almost 80% of all of our care delivery.
  - Those data systems were specifically created for true transparency. Rather than rely on existing measures derived primarily from financial systems, we applied the NQF SFB methodology to identify the full necessary measurement set, then began to collect missing measures (about 30 – 50% of the required measures were missing from “state of the art” existing data systems available at the start of the effort).

- About half of all outliers (cases, physicians, hospitals, etc.) initially identified through this clinically-based, purpose-specific measurement system traced back to the data system – not the underlying clinical care.
  - We used those outlier cases to systematically identify failures in the measurement system, then corrected them. Over time, this led to a very robust quality measurement system.
4. **Don't target the entire care delivery system at the start. Instead, provide financial incentives (shared savings) to care delivery groups who can build and implement such measurement systems**, then build out from that foundation. Let the financial incentives drive positive change over time, based upon successful models generated by early participants. Don't be trapped by the "lowest common denominator," in terms of data readily available within the existing system. This method will allow progress even though it is very difficult to accurately rank providers.
  5. **As specific clinical topics mature, move from voluntary to mandatory participation.**

### **References**

Hospital Compare website; Information for Professionals; Hospital Process of Care Measures, accessed 16Mar09 at

[http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals\\_tabset.asp?activeTab=1&Language=English&version=&subTab=7#POC3](http://www.hospitalcompare.hhs.gov/Hospital/Static/InformationforProfessionals_tabset.asp?activeTab=1&Language=English&version=&subTab=7#POC3)