

## Part II

### DATA AND QUALITY

#### Chapter 4

### **The Performance Measures Ball: Too Many Tunes, Too Many Dancers?**

*By Alice G. Gosfield\**

- § 4:1 Introduction
- § 4:2 Why now more than before?
- § 4:3 Who is measuring what?—Concepts
- § 4:4 —Selection
- § 4:5 —Priority Conditions
- § 4:6 —NQF
- § 4:7 — —Structure
- § 4:8 — —Process
- § 4:9 — —Outcomes
- § 4:10 —Other Measurers and Measures—JCAHO
- § 4:11 — —CMS
- § 4:12 — —Physician Measures Generally
- § 4:13 — —AHRQ
- § 4:14 — —Patient Safety Measures
- § 4:15 — —IHI System Level Measures
- § 4:16 — —And More?
- § 4:17 Uses and users: policy controversies
- § 4:18 —Measures themselves
- § 4:19 —Burden of measurement
- § 4:20 —Are they comparable?
- § 4:21 —Quality improvement
- § 4:22 Legal Issues

---

\*I would like to thank my colleague James L Reinertsen, M.D., and Mark Chassin, M.D., for their careful review and insightful comments on this chapter.

- § 4:23 Legal issues—Constructing measures and sets
- § 4:24 —Legal challenges to measures applications
- § 4:25 —Liabilities from measurement
- § 4:26 Conclusion

**KeyCite®:** Cases and other legal materials listed in KeyCite Scope can be researched through the KeyCite service on Westlaw®. Use KeyCite to check citations for form, parallel references, prior and later history, and comprehensive citator information, including citations to other decisions and secondary materials.

—*All sectors of the health care industry should support the focused development of quality measures that enhance and improve the ability to evaluate and improve health care.*\*\*

—*What gets measured, gets done.*\*\*\*

#### § 4:1 Introduction

Across the country, the health care system is engaged in wide ranging efforts to improve the quality of care Americans receive. Although in no way comprehensive, the scope of these efforts is far more extensive than one might have imagined even ten years ago. There are a host of initiatives to enhance patient safety in health care delivery.<sup>1</sup> Those pipers who can call the tune with their dollars are increasingly paying providers for higher quality performance.<sup>2</sup> More and

---

\*\*President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry, 2 (1998), <http://www.hcqualitycommission.gov/final/chap04.html>.

\*\*\*"A maxim in quality assessment is 'what gets measured, gets done.'" McGlynn, "Six Challenges in Measuring the Quality of Health Care," *Health Affairs*, 7-21 (May/June 1997).

##### [Section 4:1]

<sup>1</sup>See, e.g., <http://www.leapfroggroup.com>; <http://www.p4ps.org>; and Wachter, "The End of the Beginning: Patient Safety Five Years After 'To Err is Human,' Quality of Care Patient Safety Web Exclusive," *Health Affairs* (Nov. 30, 2004), at <http://content.healthaffairs.org/cgi/content/full/hlthaff.w4.s34/DC1>.

<sup>2</sup>See deBrantes, "How Purchasers Select and Pay for Quality" in Nash, et al., *The Healthcare Quality Book*, 435-51 (Health Administration Press 2004); Gosfield, "P4P Contracting: Bold Leaps or Baby Steps?," *Patient Safety & Quality Healthcare*, 30-35 (Oct./Dec. 2004); and Gosfield,

more health care report cards<sup>3</sup> respond to the Institute of Medicine's call for more transparency<sup>4</sup> with data about health plans, physicians, hospitals, nursing homes and home health agencies.<sup>5</sup> What began in 1995 with the National Committee for Quality Assurance's ("NCQA's")<sup>6</sup> first cross-plan report card on 21 health plans has now become an annual NCQA "State of Health Care Quality" report which in 2004 set forth data regarding the quality of care provided by 262 health plans covering 69 million lives.<sup>7</sup> Today, we even have the government's own "National Health Care Quality Report" issued for the first time in 2003 by the Agency for Healthcare Research and Quality (AHRQ)<sup>8</sup>, in response to a 1998 recommendation by the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry.<sup>9</sup> All of these efforts turn on performance measurement.

Yet, for all of the apparent vitality and diversity of these initiatives, the state of the art of the measurement upon which all of these undertakings turn is still evolving. Which measures ought to be used for what? How are they constructed and then selected? Whose measures will predominate? How ought they legitimately be applied? What are their boundaries and limitations? What are their potential impacts? These are all currently subjects of major policy development. Because of the many potential uses of measurement including for payment, for network selection, and as some have speculated, in liability suits, the stakes in performance measurement are high.

---

"Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 103-82 (A. Gosfield, ed. 2004).

<sup>3</sup>Gosfield, "Health Care Report Cards: Quality in the Public's Cross Hairs," *Health Law Handbook*, 501-42 (A. Gosfield, ed. 2000).

<sup>4</sup>Corrigan, et al., eds., "Crossing the Quality Chasm" (National Academy Press, Wash. D.C., 2001).

<sup>5</sup>See <http://www.healthgrades.com>; <http://www.ncqa.org/index.asp> "Nursing Home Compare;" <http://www.medicare.gov/NHCompare/Include/DataSelection/Questions/SearchCriteria.asp?version>; and text at § 4:11.

<sup>6</sup>The National Committee for Quality Assurance.

<sup>7</sup><http://www.ncqa.org/communications/SOMC/SOHC2004.pdf>.

<sup>8</sup><http://qualitytools.ahrq.gov/qualityreport/documents/E%20ExecSumm121103qg.pdf>.

<sup>9</sup><http://www.hcqualitycommission.gov>.

Despite the unsettled nature of the policy debates, as is frequently the case, neither Congress nor the law waits for perfection or resolution. In the Medicare Prescription Drug, Improvement and Modernization Act of 2003<sup>10</sup>, among the demonstrations, pilot projects and mandated studies, performance measurement is highlighted on the public health care benefit radar screen. Most directly, Congress has called on the Institute of Medicine to evaluate and report both on: (1) leading health care performance measures in the public and private sectors; and (2) options to implement policies that align performance with payment under Medicare.<sup>11</sup> That report is expected in June 2005. It is to address the validity of measures themselves as well as the success and utility of alternative performance incentive programs. It is then expected to identify and prioritize options to implement policies that align performance with payment under Medicare.

Not done with its Congressionally mandated assignments, the Institute of Medicine is also charged with studying and evaluating the activities of Quality Improvement Organizations (“QIOs”)<sup>12</sup> under Medicare including the extent to which these entities actually improve beneficiary care, the effectiveness of their reviews and whether other entities might perform these functions more effectively.<sup>13</sup> Since a major component of their work turns on performance measures and their applications,<sup>14</sup> the report due in June 2006, is likely to raise additional issues around those aspects of the work of QIOs.

On hospital performance measurement, what began as a CMS voluntary reporting initiative Congress converted to one that was not so voluntary. In § 501, Congress established that hospitals that failed to report their performance in accordance with CMS’s measures would have their Medicare payment reduced by 0.4% annually. As a result, more than

---

<sup>10</sup>Pub. L. No. 108-173, 117 Stat. 2046 (Dec. 8, 2003).

<sup>11</sup>Section 238.

<sup>12</sup>These organizations are designated as Professional Review Organizations in 42 U.S.C.A. §§ 1320c, et seq., but in 2003, by fiat, the Secretary of HHS designated them as Quality Improvement Organizations (“QIOs”), which name apparently is seen as legitimate despite the failure of Congress to amend the statute on point.

<sup>13</sup>Section 109.

<sup>14</sup>See § 4:4.

3,900 hospitals reported data that they had not previously reported, bringing participation to 98% of Medicare hospitals.<sup>15</sup> In addition, what had been launched the year before as the “National Hospital Voluntary Reporting Initiative” is now designated “Hospital Quality Alliance: Improving Care Through Information.”

Under the new Medicare Health Care Quality Demonstration,<sup>16</sup> a five year program is established to encourage the delivery of improved quality in patient care including incentives for: (1) improved safety; (2) appropriate use of best practice guidelines; (3) reduced scientific uncertainty through examination of variations in the utilization and allocations of services “and outcomes measurement and research;” and (4) encouraged share decision making between providers and patients, and more. None of these will be reportable without performance measures.

The Medicare Care Management Performance Demonstration<sup>17</sup> is a pay for performance program for physicians to meet the needs of eligible beneficiaries through the adoption and use of health information technology and “evidence-based outcomes measures.” To participate, physicians have to use information technology to manage their patients and also have to report electronically on “clinical quality and outcomes measures” established by the Secretary.

Similarly, demonstrations in home health,<sup>18</sup> adult day care,<sup>19</sup> chronic care improvement,<sup>20</sup> and the chronically ill Medicare beneficiary research, data, and demonstration strategy<sup>21</sup> all entail performance measures to some degree. With respect to the underpinnings of the actual care delivery projects, two additional sections also implicate performance measurement, namely the establishment of the Commission on Systemic Interoperability addressing how vital health

---

<sup>15</sup><http://www.medicare.gov/Hospital/Home.asp?version=alternate&browser=IE%7C6%7CWinXP&language=English&defaultstatus=1&pagelist=Home>.

<sup>16</sup>Section 646.

<sup>17</sup>Section 649.

<sup>18</sup>Section 702.

<sup>19</sup>Section 703.

<sup>20</sup>Section 721.

<sup>21</sup>Section 723.

care data systems which include performance measurement relevant data, among other things, can share information in a useful and efficient manner<sup>22</sup> and research on outcomes of health care items and services which will measure what happens to patients who are treated with specific goods and services.<sup>23</sup>

Against this backdrop of clearly intensified focus on performance measurement as a means to improve health care quality this chapter will: (1) review why such intensity now; (2) clarify what performance measures are, as well as what we are measuring and who is defining that playing field; (3) elucidate the major policy debates and controversies surrounding this movement; and (4) set forth some developing legal issues and quandaries performance measurement poses.

#### § 4:2 Why now more than before?

A raft of commentators trace the inception of health care performance measurement to Florence Nightingale and her efforts to quantify hospital infection rates, and then, before World War I to Ernest Codman and his emphasis on end results evaluation—what happens to patients who are treated.<sup>1</sup> While both are undeniably historically influential on other grounds, and clearly visionary, based on the subsequent desultory evolution of the field, it can hardly be said that either of them captured the hearts and minds of the health care system. The modern era of American performance measurement in health care, others note, goes back to the pioneering work of John Wennberg in the 1970s and on to today, identifying unexplained variation in the rate of tonsillectomies among residents of Vermont towns, now ad-

---

<sup>22</sup>Section 1012.

<sup>23</sup>Section 1013.

#### [Section 4:2]

<sup>1</sup>*E.g.*, Millenson, “Demanding Medical Excellence,” 142-47 (University of Chicago Press 1997); Moskowitz, “Ranking Hospitals and Physicians: The Use and Misuse of Performance Data,” 16 (Faulkner and Gray 1994); and Dans, “Caveat Doctor: How to Analyze Claims-Based Report Cards,” 25 *J. Quality Improvement* (Jan. 1999).

addressing the national variation in treatment of Medicare beneficiaries for disparate conditions.<sup>2</sup>

In public law, the advent of the PSRO program in the first major overhaul of Medicare in 1972, established a program, originally designed to apply clinically relevant “norms, criteria and standards” to the delivery of care in the Medicare and Medicaid programs. The goals were to reduce unnecessary services, identify medical care delivery problems and attempt to develop corrective measures where possible.<sup>3</sup> The PSRO program came to fruition in an era of hospital cost report based reimbursement and physician fee for service payment. By 1982, subjected to stinging criticism, the program was repealed and replaced with a new version which was designated “Utilization and Quality Control Peer Review Organizations,” colloquially known as “PROs.”<sup>4</sup>

Then, even though focused far more around the incentives to hospitals to underserve patients as a result of the new Medicare prospective payment system, the law reenacted the provisions on norms, criteria and standards. Regulations published in 1985 gave PROs access to hospital data, established disclosure policies and created a program to disseminate information about hospital performance gleaned from PRO review and data access. First published in 1986, but only after a Freedom of Information Act request from the New York Times, for a while the PRO program published annual reports on hospital mortality rates. But the reports created a firestorm of controversy and the “death lists” were eventually terminated.<sup>5</sup> Still, around the same time several states—notably Pennsylvania, New York, Florida and California—began to publish their own hospital mortality and

---

<sup>2</sup>See Wennberg, “Dealing with Medical Practice Variations: A Proposal for Action,” *Health Affairs*, 6-32 (1982), and “Wrestling with Variation: An Interview with Jack Wennberg, Variations Revisited, A Supplement to Health Affairs,” Web Exclusive Collection, 73-80 (2004).

<sup>3</sup>See Gosfield, “PSROs: The Law and the Health Consumer,” 29-66 (Ballinger Press 1975), and Gosfield, “Medical Necessity in Medicare and Medicaid: The Implications of Professional Standards Review Organizations,” 51 *Temple Law Quarterly* 229 (1978).

<sup>4</sup>For a review of the differences between the programs, see Gosfield, “PROs: A Case Study in Utilization Management and Quality Assurance,” *Health Law Handbook*, 361-97 (A. Gosfield, ed. 1989).

<sup>5</sup>See Moskowitz, “Ranking Hospitals and Physicians: The Use and Misuse of Performance Data,” 75-93 (Faulkner and Gray 1994).

morbidity scorecards, most around coronary artery bypass graft procedures and/or acute myocardial infarction.<sup>6</sup>

Contemporaneously, Congress was also continuing to address quality and utilization issues with its 1989 creation of the Agency for Health Care Policy and Research (“AHCPR”) “to enhance the quality, appropriateness and effectiveness of health care services.”<sup>7</sup> The Agency was to develop, review and update, clinical practice guidelines, performance measures and medical review criteria.<sup>8</sup> The Agency ended up paying far more attention to the development of guidelines than performance measures and to fighting annual budgetary battles before its demise. Performance measures, per se, barely came to light. AHCPR’s attention there was mostly aimed at how to take guidelines and apply them in hospital based quality review.<sup>9</sup>

While all of this activity demonstrated legislative and policy awareness of a useful role for some measurement of health care performance, most commentators agree that the rise of managed care and value purchasing dramatically propelled interest in performance data for comparative purposes.<sup>10</sup> The first major initiative of this kind came in HEDIS<sup>®</sup>, the Health Plan Employer Data and Information Set, which had been launched in 1989 by a group of HMOs and another select group of four large employers which were seeking to decide which plans among those in the market they would offer to their employees. The HMOs were troubled by the administrative burden of providing data in response to variable employer requests for proposals to serve their populations. The employers were bothered by the inability to compare the plans in any meaningful way. HEDIS<sup>®</sup>

---

<sup>6</sup>Gormley, “Assessing Health Care Report Cards,” 8 J. Pub Admin. Research and Theory 325 (July 1998) and Gosfield, “Health Care Report Cards: Quality in the Public’s Cross Hairs,” *Health Law Handbook*, 501-42 (A. Gosfield, ed. 2000).

<sup>7</sup>42 U.S.C.A. § 299.

<sup>8</sup>See Gosfield, “Clinical Practice Guidelines and The Law: Applications and Implications,” *Health Law Handbook*, 65-99 (A. Gosfield, ed. 1994).

<sup>9</sup>Palmer, et al., “Understanding and Choosing Clinical Performance Measures for Quality Improvement: Development of a Typology,” AHCPR, Contract No. 282-92-0038 (Jan. 31, 1995).

<sup>10</sup>See Gosfield, “Measuring Performance and Quality: The State of The Art and Legal Concerns,” *Health Law Handbook*, 31-68 (A. Gosfield, ed. 1995).

developed common measures to be applied across the plans so the employers could make legitimate comparisons.

The first HEDIS<sup>®</sup> data was issued in 1991. When the project was turned over to NCQA two years later, the purposes were also announced to include assistance to health plans to improve the quality of health care services on the basis of performance characteristics that would identify whether gaps existed between actual and expected performance.<sup>11</sup> As NCQA convened a Committee on Performance Measurement (“CPM”) to address what measures should be used for this purpose the challenges were significant,<sup>12</sup> and the science of comparative quality hardly mature.<sup>13</sup>

HEDIS<sup>®</sup> has evolved considerably from its earliest iterations and has proven a significant and vital component of the health plan comparative performance movement which has flourished in recent years.<sup>14</sup> But still, the explosion of interest in widespread performance measurement did not really find its accelerant until the publication of the President’s Advisory Commission Report on Consumer Protection and Quality in the Health Care Industry combined with the continuing impact of the Institute of Medicine’s study “Crossing the Quality Chasm.”<sup>15</sup>

The President’s Commission was convened during the high water moment of managed care backlash. State anti-managed care legislation was rampant across the country, and continuing congressional debate over “patients’ rights” confronted the ability of patients to sue their HMOs when they were denied care they wanted and were harmed as a result. The Commission was to advise President Clinton on changes in the health care system and make recommendations on steps to both assure quality and value, and protect patients and workers and their rights. The final report,

---

<sup>11</sup>See Gosfield, “Guide to Key Legal Issues in Managed Care Quality,” 203-06 (Faulkner and Gray 1996).

<sup>12</sup>McGlynn, “Six Challenges in Measuring the Quality of Health Care,” *Health Affairs*, 7-21 (May/June 1997).

<sup>13</sup>Lohr, “How Do We Measure Quality?,” *Health Affairs*, 22-25 (May/June 1997).

<sup>14</sup>See Gosfield, “Health Care Report Cards: Quality in the Public’s Cross Hairs,” *Health Law Handbook*, 501-42 (A. Gosfield, ed. 2000).

<sup>15</sup>Corrigan, et al., eds., “Crossing the Quality Chasm” (National Academy Press, Wash. D.C., 2001).

“Quality First: Better Health Care for All Americans” in March 1998, recommended a major initiative to develop a new framework and capacity for quality measurement and reporting.

Quality measurement requirements and the detailed specifications for measurement and reporting should be determined, the report recommended, through a stable and predictable mechanism suggested to be a Forum for Health Care Quality Measurement and Reporting, a private entity to involve stakeholders and represent the interests of users of information on quality. Steps should be taken to ensure that comparative information on health care quality is valid, reliable, comprehensible, and widely available in the public domain, the Commission said. The report anticipated that measures would be used to support marketplace decision-making, oversight efforts, for public health initiatives and policymaking, as well as for internal quality improvement efforts.

At that time the Commission noted that the Foundation for Accountability (“FACCT”)<sup>16</sup> had endorsed sets of health care quality measures focused on diabetes, depression, asthma, and breast cancer, as well as measurement sets that addressed health risk behaviors, plan satisfaction and the health status of the elderly. The JCAHO<sup>17</sup> had developed 42 health care quality measures and endorsed those of other organizations. NCQA’s HEDIS<sup>®</sup> program entailed more than 75 performance measures including technical quality and patient satisfaction. AHCPR, cited as a sponsor of much of the basic research in quality measurement to that point, supported the development of data collection instruments and more than 75 consumer satisfaction measures. In its CONQUEST program, the Agency was noted to have identified more than 53 separate measurement sets containing more than 1,100 clinical quality measures.

Still, the Commission found, the multiple measures did not meet the needs of potential users and the lack of widely agreed upon priorities and standards for quality measurement had been a source of both frustration and inefficiency. Because of the inability to assess the results, processes, and

---

<sup>16</sup><http://www.facct.org>.

<sup>17</sup>Joint Commission for Accreditation of Healthcare Organizations.

subjective experiences associated with every interaction with the health system or even the treatment of every disease, the Commission called for “core sets of measures for each sector of the health care industry—for reporting in a standardized way.”<sup>18</sup> To do this more efficiently, measurement priorities would have to be identified. Rather than reinvent measures, the suggestion was to obtain already created measures to include in core measurement sets.

These measures would have to be evaluated as to scientific soundness, importance of the quality concern addressed, relevance to users, potential to foster improvement in health-status or well-being, evidence basis, interpretability, actionability (the degree to which steps can be taken to address the concern), feasibility and ease and cost-effectiveness of measurement.<sup>19</sup> Then, for purposes of combining them into “sets” to be applied to each sector of health care, additional issues would have to be addressed including the comprehensiveness of the set, the representativeness of the range of measure sets and the measurement burden the set might impose.<sup>20</sup> Standardization of measurement specifications would be important for efficiency and comparability. There would have to be a way to know that providers and other organizations were using the same types of data, addressing the same segments of the population and adjusting for risk in the same ways in their calculations, to allow for confidence in comparability.

The next chapter of the report called for creating a public entity Advisory Council and a private entity as a Forum on Quality Measurement and Reporting. The Advisory Council exists, in essence, in AHRQ, which is lodged in the Department of Health and Human Services. By 1999, the Forum had also come into existence under the rubric “the National Quality Forum” (“NQF”).<sup>21</sup> As the Forum began its operations, gathering stakeholders as members, developing a

---

<sup>18</sup>President’s Advisory Commission on Consumer Protection and Quality in the Health Care Industry, Ch. 4, p.5 (1998), <http://www.hcqualitycommission.gov/final/chap04.html>.

<sup>19</sup>See discussion at § 4:4.

<sup>20</sup>President’s Advisory Commission on Consumer Protection and Quality in the Health Care Industry, Ch. 4, p.8 (1998), <http://www.hcqualitycommission.gov/final/chap04.html>.

<sup>21</sup><http://www.qualityforum.org>. See discussion at § 4:5.

governance process and calling together a Framework Board to set an agenda for its activities, the Institute of Medicine had published in 1999 its seminal work “To Err Is Human” which created the current attention to patient safety; and then in 2001 it published “Crossing The Quality Chasm” which picked up many of the Commission’s themes.

Primary among the IOM recommendations was the development of an explicitly articulated set of aims to drive quality policy, programs and delivery of care in America in the 21st century. Although not originally branded, these aims have been slightly reordered to create the memorable acronym of the “STEEEP” values. Care should be safe, timely, efficient, equitable, effective and patient-centered. Ten additional rules to redesign the health care system were promulgated, among which were at least three which were directly relevant to the performance measurement issues the Commission had identified: (1) evidence-based decisionmaking; (2) the need for transparency; and (3) shared knowledge and the free flow of information.<sup>22</sup>

On the theme of prioritization, the IOM called on AHRQ, with the NQF, to identify no fewer than 15 priority conditions, taking into account frequency of occurrence, health burden, and resource use. Most of the attention, the IOM said, should be devoted to chronic conditions which the IOM said would include cancer, diabetes, emphysema, high cholesterol, HIV/AIDS, hypertension, ischemic heart disease, stroke, arthritis, asthma, gall bladder disease, stomach ulcers, back problems, Alzheimer’s and other dementias, and depression and anxiety disorders.

So, by 2001, articulated by publicly trustworthy bodies, in directive expositions, with practical guidance and a reinvigorated sense of urgency, we now had a national strategy to confront the longstanding but largely unconsummated promise in quantifying health care quality to improve it. These two reports brought a new definition of the overarching values to drive the development of increased knowledge about, infrastructure to support, and capacity to directly improve the quality of health care delivered to

---

<sup>22</sup>The others were: (1) care based on continuous healing relationships; (2) customization based on patient needs and values; (3) the patient as the source of control; (4) safety as a system property; (5) anticipation of needs; (6) continuous decrease in waste; and (7) cooperation among clinicians.

Americans. And all of it would depend to some degree on credible performance measurement. The invitation to the performance measures ball had been issued. What dances would fill whose dance cards?

### § 4:3 Who is measuring what?—Concepts

Measures themselves have been variously called “clinical indicators,” “performance measures,” “medical review criteria” and “quality assessment criteria.” Although these terms are frequently interchanged, in advising AHCPR on the implementation of its then newly enacted authority to develop clinical practice guidelines, medical review criteria, standards of quality and performance measures, the IOM took the position that medical review criteria were “systematically developed statements that can be used to assess the appropriateness of specific healthcare decisions, services and outcomes.”<sup>1</sup> Struggling with the meaning of “performance measures,” the committee characterized the definition they did publish as “provisional” and stated that “measures” are “methods or instruments to estimate or monitor the extent to which the actions of a healthcare practitioner or provider conform to practice guidelines, medical review criteria, or standards of quality.”<sup>2</sup> Two years later in its major report on “Guidelines for Clinical Practice”<sup>3</sup> scant attention was paid to performance measures.

The purpose of using measures is to identify meaningful aspects of health care delivery that reveal something significant when analyzed. “You cannot improve what you cannot measure” is a truism in quality policy. The dimensions of health care quality that can be measured are myriad. For many years the classic characterization of the three aspects

---

#### [Section 4:3]

<sup>1</sup>Field and Lohr, “Clinical Practice Guidelines: Directions for a New Program,” 8 (National Academy Press 1990).

<sup>2</sup>Field and Lohr, “Clinical Practice Guidelines: Directions for a New Program,” 8 (National Academy Press 1990).

<sup>3</sup>Field and Lohr, eds, “Guidelines for Clinical Practice: From Development to Use” (National Academy Press 1992).

of care to review were structure, process and outcomes.<sup>4</sup> “Structure” is the environment in which care is delivered including licensure, conformity with health and safety codes and the like. For a group practice, a structure measure might be the percent of board certified physicians in the group. For a nursing home it might be the nurse to patient staff ratio. “Process” is what care is actually delivered to the patient. For hospital care a process measure could be the percentage of patients with a diagnosis of acute myocardial infarction who receive beta blockers on discharge or have a cardiac intervention within 120 minutes of admission. “Outcomes” or “end results” are what actually happened to the patient as a result of care. There have been disputes over whether a measure such as percentage of patients with a blood pressure less than 130/80 mm hg is, in fact, a real outcome measure or is an intermediate process measure since a true outcome would refer to health status-related indicators such as whether the patient’s pain subsided or full function was regained.<sup>5</sup> Now that “Crossing The Quality Chasm” has enumerated the six aims for improvement—safety, effectiveness, patient centeredness, timeliness, efficiency and equity—these are dimensions of quality that are susceptible to evaluation. Technical performance, such as the skill level of the surgeon in terms of rate of complications, can also be quantified, (although complication rates also reflect other system factors such as pre-operative preparation, antibiotic administration, and many other complex elements of a surgical experience).

The patient’s perception of the care received and communication with him have steadily advanced as critical measures of quality. In the first iterations of HEDIS<sup>®</sup>, the patient’s experience with the health plan in terms of a member’s access to service—whether getting information from the plan itself or getting appointments with physicians—was one of the earliest aspects of care reported to the public. The Foundation for Accountability came into existence to influence more performance measurement state-

---

<sup>4</sup>See Lloyd, “The Search For a Few Good Indicators,” in Nash, et al., *The Healthcare Quality Book*, 89-116 (Health Administration Press 2004).

<sup>5</sup>Wiszewainski, “Basic Concepts of Health Care Quality,” in Nash, et al., *The Healthcare Quality Book*, 25-41 (Health Administration Press 2004).

ments in terms of the patient's perceptions of care that would permit individuals to make decisions about where to seek care and permit group purchasers to negotiate for and evaluate the attainment of specific goals.<sup>6</sup> They said three kinds of information could best do this: information on results of care, whether processes follow best practices, and whether the patient's experience of care is meeting their needs and expectations. The science of measuring the patient's experience of care has itself advanced both in the health plan setting with NCQA's Consumer Assessment of Health Plans ("CAHPS<sup>®</sup>") and the hospital version of it which is being developed with the support of AHRQ.<sup>7</sup>

For certain purposes, as the President's Commission had pointed out, minimum data sets, core measures, and measure sets would be important. By specifying related significant measures associated with specific processes of care (*e.g.*, treatment of heart attack, congestive heart failure, diabetes) bundled together, a better view of the treatment of the patient would emerge. When data for a set of measures is cumulated, a better assessment of the treatment by that hospital, health plan, nursing home or physician can be made. This approach has dominated many of the most recent measure initiatives.

#### § 4:4 Who is measuring what?—Selection

In the President's Advisory Commission report, the factors for selecting and evaluating measures were identified as critical. Similar to the eight desirable attributes for clinical practice guidelines that had been identified in its 1990 study<sup>1</sup>, in 1992, the IOM identified eight desirable attributes of medical review criteria. Since their definition conforms to today's definition of performance measures, the attributes are relevant here. Medical review criteria should demonstrate: (1) sensitivity (it is likely that a case will be

---

<sup>6</sup>Lansky, "Measuring What Matters to the Public," *Health Affairs*, 40-41 (July/Aug. 1998).

<sup>7</sup>See Edgman-Levitan, "Measuring and Improving Patient Experiences of Care," in Nash, et al., *The Healthcare Quality Book*, 183-212 (Health Administration Press 2004).

#### [Section 4:4]

<sup>1</sup>Field and Lohr, "Clinical Practice Guidelines: Directions for a New Program," 8 (National Academy Press 1990).

identified as deficient given that it really is deficient); (2) specificity—it is highly likely that the criteria will identify truly good care as such; (3) patient responsiveness—the criteria specifically identify a role for patient preferences in the process for using them; (4) readability—they are presented in language and formats that can be read and understood by non-physicians including patients, as well as practitioners; (5) minimum obtrusiveness—the process for applying them minimizes inappropriate direct interaction with and burdens on the treating practitioner or patient; (6) feasibility—the information required for review can be obtained easily, the criteria are easy to apply and are accompanied by explicit instructions for their application and scoring; (7) computer compatibility—they are straightforward enough to be transformed readily into computer based protocols and formats; (8) appeals criteria—they provide explicit guidance about the considerations to be taken into account when adverse review decisions are appealed by professionals or patients.<sup>2</sup>

In 1997, the three critical issues which drove the development of HEDIS<sup>®</sup> measures were: (1) the relevance of the measure to actually enhancing the health of the population under review and whether the measure itself will permit action to be taken upon it; (2) the scientific soundness of the measure including its reliability (repeated measurement produces the same results), validity (the measure really reflects the quality of care delivered), and adjustability (factors other than quality are accounted for in the final score) are important; and (3) The feasibility of the measurement as a practical matter.<sup>3</sup> To the extent that it is difficult to find data in a manageably retrievable way to evaluate performance, the measure will be less useful. In other instances, the problem is that for measurement to have meaning, there has to be enough instances of the studied behavior to be

---

<sup>2</sup>Field and Lohr, eds., “Guidelines for Clinical Practice: From Development to Use” (National Academy Press 1992).

<sup>3</sup>See McGlynn, “Six Challenges in Measuring the Quality of Health Care,” *Health Affairs*, 7-21 (May/June 1997). “Validity for process and structure measures means they have a proven relationship to outcomes we care about.” For outcome measures to be “valid,” they must bear a proven relationship to processes we can modify to affect the outcome.” Mark Chassin, M.D., *Personal Communication* (Jan. 3, 2005). See also Chassin, “Quality of Care: Time to Act,” 1991 JAMA 3472-73 (1991).

significant. If a specific physician group has few patients who are subscribers to a specific health plan, the data that health plan might recover regarding the physicians' performance may not be meaningful.

As concern has intensified over problems of misuse, overuse and underuse in American healthcare<sup>4</sup>, perhaps the most critical aspect of performance measurement today is the need to ground measures as much as possible in science. We need to be sure that what we are measuring reflects what science says is the right intervention for the patient's clinical needs. As the science of quality measurement has improved, many have agreed that both process and outcome measures are appropriate and neither is best.<sup>5</sup> On the other hand, for a good number of years in quality policy, the Holy Grail for evaluation was thought to be outcomes data.<sup>6</sup> Speaking principally about health plan measures, "We prefer to measure health outcomes because they are what people really care about, they are comprehensible, they aggregate the effects of all the things plans do for a condition, and they leave plans free to determine for themselves the best things to do."<sup>7</sup> But the ability to generate any substantial number of significant, valid outcomes measures proved elusive.

By 1998, David Eddy had identified five basic problems in outcomes measurement for plans that are also relevant to hospitals and health systems: (1) probabalism; (2) rarity; (3) delay; (4) weak control; and (5) confounding.<sup>8</sup> In order, these stand for the following propositions: outcomes do not always occur when a plan does the right thing and can occur even

---

<sup>4</sup>Chassin, Galvin, et al., "The Urgent Need to Improve Health Care Quality," 280 JAMA 1000 (Sept. 16, 1998).

<sup>5</sup>See Lloyd, "The Search For a Few Good Indicators," in Nash, et al., *The Healthcare Quality Book*, 89-116 (Health Administration Press 2004), with Wiszewainski, "Basic Concepts of Health Care Quality," in Nash, et al., *The Healthcare Quality Book*, 25-41 (Health Administration Press 2004); Lohr, "How Do We Measure Quality?," *Health Affairs*, 22-25 (May/June 1997); and McGlynn, "Six Challenges in Measuring the Quality of Health Care," *Health Affairs*, 7-21 (May/June 1997).

<sup>6</sup>See Moskowitz, "Ranking Hospitals and Physicians: The Use and Misuse of Performance Data," 16 (Faulkner and Gray 1994).

<sup>7</sup>Eddy, "Performance Measurement: Problems and Solutions," *Health Affairs*, 11 (July/Aug. 1998).

<sup>8</sup>Eddy, "Performance Measurement: Problems and Solutions," *Health Affairs*, 11 (July/Aug. 1998).

when a plan does the wrong thing. The most important things we would want to measure occur relatively rarely across a single plan's population. It takes many years, five to ten years of longitudinal information, to get to survival rates and true outcomes. The extent to which a plan can actually control what produces an outcome is questionable given the multi-factorial influences on outcomes. In large networks, a single plan may have even less influence over the delivery of care than something else does. And, finally, the appropriate level of clinical detail at which the analysis ought be focused is a problem related to definitional problems at the core of discerning what is being measured.

These inherent problems, Dr. Eddy noted, are further complicated by implementation dilemmas that include inadequate information systems, too many measurers and measures, health plan complexity and a lack of funding. "Today's [outcomes] measures tend to be blunt, expensive, incomplete and distorting, and they can easily be inaccurate and misleading."<sup>9</sup> The good news is there has been evolution since 1998.

There is a broadly agreed view that meaningful measures come from an evidence base. Clinical practice guidelines reflecting scientific evidence are strongly related to and provide a basis to develop performance measures. The distinctions between CPGs and measures, though, can be confounding. "Clinical practice guidelines" as defined by the IOM to AHCPR<sup>10</sup> are to be used in the moment of treating the patient. "Performance indicators" or measures, on the other hand, are intended to provide information to help insure accountability for clinical performance and identify opportunities for improvement. Some have argued that because of the difference in their functions, there are tensions between them, in that CPGs reflect "ideal care" at the patient level, whereas performance measures are intended to encompass a population of patients at various stages of

---

<sup>9</sup>Eddy, "Performance Measurement: Problems and Solutions," *Health Affairs*, 11, 15 (July/Aug. 1998).

<sup>10</sup>Field and Lohr, "Clinical Practice Guidelines: Directions for a New Program," 8 (National Academy Press 1990).

disease.<sup>11</sup> Interestingly, in focusing on the fact that medical review criteria and related tools (like performance measures) emphasize the evaluation of healthcare decisions, actions and outcomes, the 1992 IOM committee noted that one committee member strongly objected to the distinction between guidelines and review criteria taking the position that “there should not be one iota of difference between a good guideline intended for [practitioners] and a medical review criteria intended to assess care; they are different uses of the same clinical statement.”<sup>12</sup> Some believe that the increasingly widespread use of performance indicators (“PIs”), which are less stringent than idealized clinical practice guidelines, risks lowering the standard of clinical care. “Because [performance measures] allow healthcare organizations to demonstrate accountability to a variety of audiences, they must be based on evidence for which the greatest consensus exists.”<sup>13</sup> The issue of consensus is a significant aspect of the current performance measures landscape.

#### § 4:5 Who is measuring what?—Priority Conditions

Consensus on prioritization among the universe of conditions, treatments, and dimensions of quality that might be the subject of measurement was a critical aspect of the President’s Advisory Commission report. In “Crossing the Quality Chasm,” the IOM recommended that AHRQ consider prioritizing its work in accordance with the list of conditions identified through the Medical Expenditure Panel Survey, a nationally representative household survey of healthcare use, expenditures, sources of payment and insurance coverage. These are the same conditions that had been identified, from the same source, as in the President’s Commission Report. It was anticipated that the number of prior-

---

<sup>11</sup>O’Malley, Clancey, et al., “Clinical Practice Guidelines and Performance Indicators as Related-But Often Misunderstood-Tools,” Joint Commission Resources, at <http://www.jcrinc.com/subscriber/printview.asp?durki=8019> (accessed Oct. 27, 2004).

<sup>12</sup>Field and Lohr, “Clinical Practice Guidelines: Directions for a New Program,” 27 (National Academy Press 1990).

<sup>13</sup>Field and Lohr, “Clinical Practice Guidelines: Directions for a New Program,” 27 (National Academy Press 1990).

ity conditions identified should grow over time to eventually cover the majority of care provided to patients.<sup>1</sup>

The study made the case for the value of organizing around priority conditions. Such an agreed upon focus could align the efforts of diverse participants in the healthcare system, offer meaningful levels of organization to patients, and provide a starting point on which healthcare professionals and organizations can direct their energies. Defining care processes around specific conditions would establish a suitable level of focus for significant quality improvement in healthcare.

The report explicitly acknowledged that most priority conditions would be strongly related to chronic conditions and identified the fact that the current healthcare system is organized around acute care needs which does not facilitate the flow of information over time, offers little recognition or reward for coordinating care, and pays mainly for face to face visits, not for information or for reassurance at other times.<sup>2</sup> Most significantly for the issue of performance measurement itself, the priority conditions were seen to improve the feasibility of quality measurement by offering a framework for the development of standards to guide the necessary data collection. The criteria for identifying priority conditions were said to include prevalence, burden of illness, cost, variability in practice, and a potential to improve outcomes or reduce costs.<sup>3</sup>

But as part of the broader scope of its quality initiative, the IOM convened yet another committee and issued a still further report addressing selection of priority conditions.<sup>4</sup> In

---

**[Section 4:5]**

<sup>1</sup>Corrigan, et al., eds., *Crossing the Quality Chasm* (National Academy Press, Wash. D.C., 2001).

<sup>2</sup>Corrigan, et al., eds., “Crossing the Quality Chasm,” 93 (National Academy Press, Wash. D.C., 2001).

<sup>3</sup>Corrigan, et al., eds., “Crossing the Quality Chasm,” 103 (National Academy Press, Wash. D.C., 2001).

<sup>4</sup>The IOM had issued three other reports on priorities for action including priorities for technology assessment (1992), clinical practice guidelines (1992), and research funding (1998).

2003 the IOM recommended<sup>5</sup> three criteria to establish national priorities: (1) impact—the extent of the burden in terms of disability, mortality and economic costs, including effects on patients, families, communities and societies; (2) improvability—not only the extent of the gap between current performance and evidence based best practice but also the likelihood that the gap can be closed through change and there is opportunity to make dramatic improvements in the six STEEEP aims; and (3) inclusiveness—the sweep of the relevance to a broad sector of the population, the ability to generalize associated quality improvement strategies to many types of conditions and illnesses, in other words how representative the condition is of other conditions that should also be improved, and the breadth of change effected through such strategies across a range of health care settings and providers.<sup>6</sup> While chronic conditions were identified as strong candidates for inclusion, despite the earlier position in “Crossing The Quality Chasm” that chronic conditions should be the primary focus of identifying the priority conditions to transform the health care system, the committee’s charge was to look beyond chronic conditions. Still, the central principle was that by constructing a clearly articulated list reflecting the STEEEP values, broad application would emerge in part based on consensus. Consequently, much of the attention in performance measurement over the short term would reflect these conditions. As it is unfolding, however, the priority list is not all that will or has emerged as topics of performance measurement, even as considerable consensus has focused around the need to first address these conditions.

---

<sup>5</sup>Adams and Corrigan, eds., “Priority Areas for National Action: Transforming Healthcare Quality” (National Academy Press 2003).

<sup>6</sup>The final priority list included both crosscutting and condition specific priorities: care coordination, self-management health literacy, asthma, cancer screening, children with special health needs, diabetes, end of life with advanced organ system failure with a focus on congestive heart failure and chronic obstructive pulmonary disease, frailty associated with old age, hypertension, immunization, ischemic heart disease, major depression, medication management, nosocomial infections, pain control in advanced cancer, severe and persistent mental illness, stroke, tobacco dependence and obesity.

**§ 4:6 Who is measuring what?—NQF**

The National Quality Forum was created in response to the President's Advisory Commission. It has been organized specifically to comply with the criteria set forth in the National Technology Transfer and Advancement Act of 1995<sup>1</sup> which requires all federal agencies and departments to use technical standards that are developed or adopted "by voluntary consensus standard bodies, using such technical standards as a means to carry out policy objectives or activities determined by the agencies and departments."<sup>2</sup> As further set forth in OMB Circular A-119, there are surprisingly detailed specifications for how voluntary consensus standards will be developed and how federal agencies may participate with bodies that develop them. The significance is that the work of the NQF, at which the Center for Medicare and Medicaid Services ("CMS") sits as a member, will directly impact on the federal programs that CMS regulates.

**§ 4:7 Who is measuring what?—NQF—Structure**

The structure of NQF is designed around the need to develop consensus decisions. Therefore, the current 250 entities membership is voluntary, self-generated and organized around stakeholder status. Four Membership Councils reflect the constituencies which produce the consensus: (1) consumer and patient groups; (2) purchasers; (3) providers and health plans; and (4) research and quality improvement organizations. Membership in this public-private partnership includes national, state, regional and local groups representing consumers, public and private purchasers, employers, healthcare professionals, provider organizations, health plans, accrediting bodies, labor unions, supporting industries, and business coalitions, consultants and organizations involved in healthcare research for quality improvement. Membership is purely voluntary and self-selected. Those organizations that want to pay dues and participate may do so. There are no individual members. NQF is governed by a 23-member board of directors, including

---

**[Section 4:6]**

<sup>1</sup>Pub. L. No. 104-113, 1996 HR 2196, 110 Stat 775.

<sup>2</sup>15 U.S.C.A § 272(d).

representatives from CMS and AHRQ. All other members of the board are elected by the Forum voting through the Member Councils. The Councils are the vehicles through which consensus is achieved. In addition, there is a Strategic Advisory Council of experts who are called on in certain instances.

To begin its work, the NQF convened a Strategic Framework Board to propose a conceptual framework for the ensuing work. This nine member group produced recommendations which yielded 17 NQF consensus endorsed principles and statements of policy that have guided the work of the NQF going forward. Without reiterating all of those principles here, in general, the Framework Board recommended that NQF's goals be consistent with the six aims of the Board<sup>1</sup> and should relate to clinical conditions that are prevalent or have a high risk of disability, suffering, or death or that address cross cutting issues not specific to a single clinical condition. The goals should be based on evidence or expert opinions that effective clinical care or system and process improvement strategies exist and are supported by expert groups and compelling to relevant constituents. The Framework Board endorsed the development of "a parsimonious common set of quality measures that is incrementally improved based on feedback" from users.<sup>2</sup> Measures themselves to be included in the common set should collect data once, as close to their source as possible, and so that they can be combined, analyzed and reported to serve a wide variety of purposes.

Beyond pure performance measures issues, the Framework Board identified five specific strategic priority areas to: (1) educate the public, ensure that reports are compelling and useful and support decisionmaking; (2) define and develop processes for timely delivery of performance reports; (3) adopt a policy that purchasers and payors should require providers to routinely and publicly report performance on a common set of measures; (4) develop, pilot test and imple-

---

**[Section 4:7]**

<sup>1</sup>Which are just slightly different from the STEEEP aims by substituting "beneficial" for "effective."

<sup>2</sup>NQF, "A National Framework for Healthcare Quality Measurement and Reporting," 4 (2002). All NQF documents are available at <http://www.qualityforum.org>.

ment a strategy to evaluate the impact of measurement and reporting; and (5) remove barriers and negative incentives and establish rewards for quality performance. Through these strategic priority areas, NQF has begun to move into arenas that go beyond identification and endorsement of common measurement sets.

To guide its work and the implementation of its strategies and products, the Framework Board adopted policy statements that the administrative burden of reporting should be reduced by eliminating redundancy. Electronic information systems to facilitate reporting should be organized around the NQF common measures and framework. Professionals should be trained to understand skills basic to quality improvement. Provider organizations and accrediting and licensure bodies should ensure that individual providers are able to utilize performance information for decisionmaking and quality improvement. Federal government funding should be available for technical assistance for those that are moving aggressively. A national research agenda conference should be convened to address a five-year healthcare quality research agenda. Other funding priorities were also identified and recommended.

#### **§ 4:8 Who is measuring what?—NQF—Process**

It is the consensus development process that casts NQF in a unique role in performance measurement as a “voluntary consensus standard setting organization” under the Technology Transfer Act. NQF has a five step process for the endorsement of performance measures (as well as other consensus processes such as the adoption of the framework board report): (1) development; (2) review; (3) member council approval; (4) Board of Directors endorsement; and (5) evaluation. Each project NQF undertakes is guided by a Steering Committee (or Review Committee) composed of individuals from each of the four critical stakeholder perspectives. NQF staff and technical advisory panels, with the ongoing input of NQF members, work with the Steering Committee to conduct an overall assessment of the state of the field in a particular topic area.

The Steering Committee recommends a set of draft measures, indicators, or practices for review, along with a rationale for proposing them. Those proposed standards are

distributed for review and comment by NQF members and non-members. All NQF consensus reports are expected to be formally “explicit about the scientific evidence and experience underlying the recommended measures or indicators, the criteria for selecting them and the rationale for recommending a particular item or approach.”<sup>1</sup> Any draft product approved by a Steering Committee or Review Committee is expected to include an appendix or attachment as commentary where members may express dissenting views or other perspectives not covered in the report.

Each member organization has the opportunity to review and comment before voting. Member councils themselves are responsible for establishing their own procedures to communicate their aggregate comments and positions on a proposed standard. Once a draft product has been provided to the member councils, it is also made available for general public review. Comments are reviewed and the drafts are then revised in a shortened period.

All members in good standing (current on dues and other invoices) are provided the opportunity to vote on any consensus project. Quorums are not required, and a straight majority is used to determine the position of the council. This puts a significant premium on voting since abstentions do not count and failure to vote does not count. If, after the first round of voting, one or more member councils is unable to obtain agreement by a majority of members of that council, the staff will attempt to resolve the matter and submit a revised draft to all councils for further consideration. Another vote is then taken. If majority agreement across all councils is not achieved after two rounds of voting but has been achieved by two councils, then the document is forwarded to the board of directors with a summary of major issues and points of disagreement.

The Board holds the ultimate authority for endorsement or reconsideration. Determinations made under this mechanism do not reflect unanimity, but consensus. The consensus process allows for a request for reconsideration for an endorsed voluntary consensus as an appeal. The appeals process is in essence a reconsideration with review by NQF

---

**[Section 4:8]**

<sup>1</sup>NQF Consensus Development Process, Version 1.7, 5 (Aug. 20, 2004).

staff and management who may consult with the project's Steering Committee, Technical Advisory panels, and any other sources as appropriate before they make a recommendation to the Board. The Board will act on all appeals within 60 days of receiving the appeals package.

The evaluation function is intended to take into account the fact that some standards may identify data collection, analysis and/or reporting issues, only after widespread use. Because it is expected that once endorsed, voluntary consensus standards would be continuously evaluated and refreshed, standing working groups are established to accommodate technical changes, to endorse standards.

The public-private nature of the NQF undertaking is also seen in their Policy on Endorsement of Proprietary Performance Measures. Because of the Technology Transfer Act, the basic strategy is for consensus standards bodies to endorse standards already developed by others. In the NQF frame of reference, this means that NQF is establishing, by selecting from other existing measures, the individual measures they endorse, but even more specifically they bundle their own measure sets. Not all measures in the market place are eligible to be considered.

In order for NQF to endorse a measure, entities have to have equal access (*e.g.*, via the web) to the specifications at no charge. All measures specifications must be fully disclosed (“open source”) meaning the data elements are fully described and disclosed, the measure algorithm is fully described and disclosed, and risk adjustment method, data elements are also described and disclosed. Providers must be permitted to use measurement specifications without charge but if a final data analysis requires obtaining, for a fee, access to a database dependent coefficient, this fact shall also be disclosed when the measure is submitted to NQF. The measure development and/or holder of the intellectual property rights for the measure has to agree to certain disclosure requirements and execute NQF's standardized agreement to this effect.



#### § 4:9 Who is measuring what?—NQF—Outcomes

Using its consensus process, after the endorsement of the Framework Board's recommendations, the NQF began with publication of an Initial Performance Measurement Set for

Hospital Care. The 39 measures are focused around acute coronary failure, heart failure, pneumonia, pregnancy/childbirth/neonatal conditions, smoking cessation, surgical complications, three pediatric conditions and four patient safety issues. NQF subsequently held a Summit on how to measure and report quality using the NQF endorsed hospital care measures.<sup>1</sup> Describing the purpose of the measures, the Summit report stated this measure set:

has been designed primarily to facilitate quality improvement through public accountability. But in their legal standing as voluntary consensus standards, the initial NQF-endorsed measures are exceptionally well positioned to serve as a foundation for single, standardized measure sets to improve hospital care through performance measurement and reporting.

The point of the hospital measure project was not to introduce any new demands for information but rather to “channel all the resources currently being wasted through scattered measurement efforts into one powerful tool that would facilitate healthcare quality improvement throughout the United States.” The Summit identified four crucial components to realize the tipping point for widespread use of the measures: achieving standardization of the measures; data collection and submission; rewarding high quality performance; and public reporting. The report noted the complexity of translating the measures themselves into practical applications of them.

Beyond hospital measures, the NQF has published consensus standards for nursing sensitive performance measurement, for nursing home care, safe practices, adult diabetes care, and serious reportable events. It has in preparation at this writing, voluntary consensus standards for ambulatory care, consumer focused measures of mammography center quality, evidence based substance abuse treatment practices, standardizing quality measures for cancer care, improving patient safety in medication use, improving patient safety through informed consent, standardizing cardiac surgery performance measures, voluntary consensus standards for home health, and for prevention and care for deep vein thrombosis. It is looking to be involved in consensus stan-

---

**[Section 4:9]**

<sup>1</sup>“Reaching the Tipping Point: Measuring and Using Quality Using the NQF Endorsed Hospital Care Measures” (2003).

dards for substance abuse treatment, credentialing simplification and standardization (with the Joint Commission and NCQA), and other activities, pending funding.

Despite the position of the IOM in its work, and even after some of its initial measures were already endorsed, in 2004 the NQF published its own 23 NQF Endorsed Priorities for Healthcare Quality Measurement and Reporting. These priorities cover the continuum of care and are organized into 2 infrastructure priorities, 5 process of care priorities, and 15 health care condition priorities. The highest priority across all of these areas is to reduce disparities in health and healthcare quality in vulnerable populations. The selected priorities are to guide NQF healthcare quality measurement, reporting and improvement activities and the endorsement of voluntary consensus standards. In undertaking new projects, NQF will place the highest priority on projects that address these priorities, although the project activities need not be limited to the priority topics.

NQF claims its priorities relate in whole or in part to the IOM 2003 priorities but are generally broader than the corresponding IOM topics. Of course, the refinement in the IOM issues (*e.g.*, not just end of life issues, as stated by NQF, but end of life associated with congestive heart failure and chronic obstructive pulmonary disease; not just pain control but pain control in advanced cancer) was quite intentional and designed to focus attention on those specific aspects of those issues as opportunities for improvement. The paradox here is that while NQF's more general priorities are for measurement and reporting alone, the IOM's more focused priorities were selected for far broader transformation of health care quality.

NQF also explicitly identifies information technology as an infrastructure issue that is a priority. Their priorities list added pneumonia and kidney disease to the topics covered in the IOM list (although they only identify pneumonia as having no grounding in the IOM topics). The effect of the differences in these priorities is not yet known; but including greater breadth may dilute the impact of the priority conditions, thereby undermining the opportunities identified by the IOM as the reason to have priority conditions enumerated in the first place.

Still further, while the NQF claims for itself a unique legal

status (and it may be right), they are not alone in addressing either this range of topics or the specifics within them. The plethora of measures is increasing and not decreasing since David Eddy's evaluation in 1998, although there is more apparent commonality (but not uniformity) among leading measure sets.

#### § 4:10 Who is measuring what?—Other Measurers and Measures—JCAHO

One of the earliest hospital related performance measurement programs was launched by the Joint Commission in 1987 as part of its agenda for change. The idea was that by incorporating performance measurement into accreditation, the decisions would be more meaningful. Its ORYX initiative was intended to allow hospitals flexibility in selecting vendors of performance measurement programs.<sup>1</sup> The program was widely criticized for lack of specificity and risk adjustment. Now the JCAHO has moved to a core measurement set which is derived largely from the PRO's sixth scope of work grouped into three sets: nine measures about acute myocardial infarction; four measures about heart failure; five measures about community acquired pneumonia; and a measure set on pregnancy and related conditions.<sup>2</sup> The IOM observed that the ORYX initiative encompasses four core measurement areas that are applicable to their selected priority areas.

Under the ORYX program, the healthcare organization chooses an accredited system vendor, selects clinical measures that cover a specific percentage of its patient population, implements data collection and submits data to their systems vendor monthly. The system vendor calculates the measures, compares them to a norm and electronically submits summarized statistics to the JCAHO on a quarterly basis. The JCAHO monitors trends and patterns and takes

---

#### [Section 4:10]

<sup>1</sup>Lawrence, "JCAHO's Oryx Initiative Links Outcomes with Accreditation," *Managed Care* (Feb. 1998), <http://www.managedcaremag.com/archives/9802/9802.oryx.shtml>.

<sup>2</sup>"A Comprehensive Review of Development and Testing for National Implementation of Hospital Core Measures," [http://www.jcaho.org/pmf/core+measures/cro\\_hos\\_cm.htm](http://www.jcaho.org/pmf/core+measures/cro_hos_cm.htm) (accessed Oct. 24, 2004).

action in terms of accreditation. The Joint Commission approves vendors and their measures for acceptance in the reporting process.

Each vendor gets approved for specific measures such as mortality rate for specific DRG, post operative surgical complication rates, average length of stay by specified DRG, and the like. Many of the measures are approved for inclusion in the core measure sets. The Joint Commission has recently agreed with CMS to align their current and future common hospital quality measures in their condition specific performance measurement sets. Hospital quality measures are included in the Joint Commission's ORYX core measures and also CMS's seventh scope of work with measures on heart attack, heart failure, pneumonia, and surgical infection prevention.<sup>3</sup>

#### § 4:11 Who is measuring what?—Other Measurers and Measures—CMS<sup>1</sup>

CMS's role as a measurer has expanded considerably. The National Voluntary Hospital Reporting Initiative (for which hospitals can suffer reduced reimbursement for not reporting),<sup>2</sup> has a 10 measure starting set, which draws on Joint Commission measures, NQF's initial hospital set and other CMS measures and addresses three principal conditions: acute myocardial infarction, heart failure and pneumonia. This is distinct from the Premier Hospital Quality Incentive Demonstration, which is a pay for performance demonstration program with its own clinical conditions (acute myocardial infarction, heart failure, pneumonia, coronary artery bypass graft, and hip and knee replacements)<sup>3</sup> and 34 measures for reporting. Among them are included all 10 indicators from the starter set of the National Voluntary

---

<sup>3</sup>JCAHO, "CMS to Work Together to Align Hospital Quality Measures," *Private Sector News*, Vol. 8, Nos. 19 and 764 (Sept. 29, 2004).

#### [Section 4:11]

<sup>1</sup>For general information on CMS's quality initiatives, see <http://www.cms.hhs.gov/quality/default.asp?>

<sup>2</sup>See text at § 4:1.

<sup>3</sup>For a description of the leading pay for performance programs and their measures, see Gosfield, "Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 103-82 (A. Gosfield, ed. 2004).

Hospital Reporting Initiative, 27 from NQF, 24 indicators from the 7th Scope of Work for QIOs, 15 indicators from the Joint Commission Core Measures, 3 proposed by the Leapfrog Group and 4 from AHRQ's patient safety measures. The complexity of the Venn diagram to display the overlaps and differences in the sources of the measures would surely dazzle. As if culling from existing measures sets were not enough, to further support its hospital quality initiatives, CMS launched the Robust Measures Project to identify "a robust and comprehensive measures set for hospital public reporting." This project will identify an expanded set of measures for public reporting, after the starter set just noted above.<sup>4</sup>

The QIOs have core measures for hospitals dealing with acute myocardial infarction, heart failure, pneumonia and preventing surgical infections.<sup>5</sup> They also are the vehicles through which CMS obtains some of the data that it makes available regarding comparative performance of home health agencies and nursing homes. (Other aspects of that data are separately reported.) These measures are far less technically clinical in their focus, but form the basis for the "compare" programs CMS has launched to help Medicare beneficiaries select nursing homes, home health agencies and health plans (for which CMS requires reporting of HEDIS<sup>®</sup> and CAHPS<sup>®</sup> measures). Nursing home measures relate to such matters as percentage of residents who have moderate to severe pain, percentages of high risk and low risk residents who have pressure sores, the percentage of residents who are more depressed or anxious, the percentage of low risk residents who lose control of their bowels or bladder and the like.<sup>6</sup> The home health measures are more akin to this approach than they are to the hospital reporting projects. The home health quality measures report on the numbers of patients who have improvement or degradation in aspects of their status that relate to improvement in getting around, meeting their needs for activities of daily living, patient medical emergencies and improvement in mental health.

---

<sup>4</sup>Robust Measure Project, Centers for Medicare and Medicaid Services Fact Sheet (Feb. 18, 2004).

<sup>5</sup><http://www.cms.hhs.gov/qio/2t.pdf>.

<sup>6</sup><http://www.medicare.gov/NHCompare/Static/Related/DataCollection.asp?dest=NAV/>.

CMS is actively measuring physician performance through the QIOs, which apply measures associated with diabetes treatment in the office, screening mammography and adult immunizations. In addition, two of its physician oriented demonstration/pilot projects incorporate performance measurement. The CMS Physician Group Practice Demonstration is a pay for performance program that was mandated by § 412 of the Benefits Improvement and Protection Act of 2000 to focus on chronically ill Medicare beneficiaries treated by six groups of at least 200 physicians each. This project draws its measures from HEDIS and the Medicare Health Care Quality Improvement Program in Medicare managed care, measuring eye exams for diabetics, HgA1c tests, lipid profiles for diabetics, mammograms, chest x-rays within three months of a diagnosis of congestive heart failure, flu and pneumonia vaccines, and office visits for patients with any of four chronic conditions.

The Doctors Office Quality Project is a different demonstration running through the three QIOs in Iowa, California and New York. Here, CMS is using performance measures that it has developed into a set with the AMA (*see below*) and NCQA. These measures have been proposed by CMS to NQF for endorsement, but they had not yet been endorsed at this writing.

#### **§ 4:12 Who is measuring what?—Other Measurers and Measures—Physician Measures Generally**

There are far fewer physician-focused measures than those regarding hospitals. Still, though, most pay-for-performance (P4P) programs are focused on physicians, including Bridges to Excellence, the Integrated Health Association initiative in California, and NCQA's Diabetes and Heart/Stroke Provider Recognition Programs, among others.<sup>1</sup> The vast majority of P4P programs, but not all, address chronic conditions, and primarily in outpatient settings. In addition, while HEDIS<sup>®</sup> has always been advertised as a health plan measurement set, in fact, significant elements of what HEDIS<sup>®</sup> measures and upon which NCQA accredits turn on the activities of

---

**[Section 4:12]**

<sup>1</sup>Gosfield, "Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 103-82 (A. Gosfield, ed. 2004).

physicians, as distinct from other elements of the system, including the health plan itself. HEDIS<sup>®</sup> measures on cholesterol and lipid levels, diabetic treatments, immunizations, screening for chlamydia, cervical cancer and breast cancer, prescription of beta blockers, administration of asthma medicine reflect activities primarily conducted by physicians. In its physician recognition programs dealing with diabetes, and heart and stroke treatment, NCQA measures both process (*e.g.*, performance of a complete lipid profile, use of aspirin, eye and foot examinations) and outcomes (*e.g.*, patient levels of blood pressure and cholesterol control). HEDIS<sup>®</sup> measures are the basis for many of the physician P4P program measures.

Still, the comparative lag in measures directly aimed at physicians stems both from the barriers to reporting that lie in physician disorganization and relative disaggregation in contrast with the quantities of data available in one hospital, for example. In addition, there has been significant resistance from physicians who have been extremely fearful of comparative reporting of any kind because they believe much of what they are measured on is not within their control; the small numbers upon which to report about them lead to distorted conclusions about them; and the data upon which the reports are drawn is flawed, particularly when it comes from health plans.<sup>2</sup>

Recognizing, however, that at this specific moment in policy evolution there is still an opportunity to shape their world, some physician organizations have put their stake in the ground on performance measurement. Notably the American College of Physicians and the American Academy of Family Physicians have joined with AHRQ and America's Health Insurance Plans ("AHIP"), the managed care lobbying group, to convene "The Leadership Project," a 60 member invitational working group, to develop a collaborative strategy to confront the particular challenges in physician performance measurement.

The AMA itself has convened the Physician Consortium for Performance Improvement which is yet another group (with some overlap) of clinical and methodological experts from more than 60 national medical specialty and state medi-

---

<sup>2</sup>Gosfield and Reinertsen, "Doing Well By Doing Good: Improving the Business Case for Quality" (2003), <http://www.uft-a.com>.

cal societies, AHRQ and CMS. Its purpose is to: (1) become the leading source organization for evidence based clinical performance measures and outcomes reporting tools for physicians; and (2) assure that all components of the medical profession have a leadership role in all national forums seeking to evaluate the quality of patient care. It will identify and develop evidence based clinical performance measures that enhance quality of patient care and that foster accountability.<sup>3</sup> The Consortium has published its own measurement sets dealing with asthma, coronary artery disease, adult diabetes, heart failure, hypertension, major depressive disorders, osteoarthritis of the knee, prenatal testing and preventive care and screening. They have also drawn on measures produced by others including the American College of Cardiology, the American Heart Association, the American Academy of Orthopedic Surgeons and others. The AMA measure sets address explicitly whether a measure is appropriate for internal use or public reporting.

#### **§ 4:13 Who is measuring what?—Other Measurers and Measures—AHRQ**

AHRQ has published its own hospital quality indicators that make use of hospital and patient administrative data. These measures focus on three aspects of quality: (1) prevention indicators which identify hospital admissions that evidence suggests could have been avoided, at least in part, through high quality out-patient care; (2) inpatient indicators reflecting quality of care inside hospitals including inpatient mortality for medical conditions and surgical procedures; and (3) patient safety indicators reflecting quality of care inside hospitals but focusing on potentially avoidable complications and iatrogenic events. Though the indicators were primarily developed for internal use by hospitals in quality improvement, AHRQ has now published a document providing “Guidance for Using the AHRQ Quality Indicators for Hospital Level Public Reporting or Payment”<sup>1</sup>

---

<sup>3</sup><http://www.ama-assn.org/ama/pub/category/2946.html>.

#### **[Section 4:13]**

<sup>1</sup>Remus and Fraser, DHHS, Rockville, M.D. AHRQ Publication No. 04-0086-EF (Aug. 2004), <http://www.qualityindicators/AHRQ.gov/documentaion.htm> More on Activities.

**§ 4:14 Who is measuring what?—Other Measurers and Measures—Patient Safety Measures**

Patient safety represents a subset of more general quality reporting, and has been a focus of still another range of performance measures. The Leapfrog Group has led the way with specific measures to be used in evaluating hospitals for purchasing decisions and payment rewards. Using a hospital reported survey on the safety dimensions of interest to Leapfrog's purchasers, the survey has always looked at the three fundamental issues of measuring: (1) whether and the extent to which hospitals have implemented computerized physician order entry programs to prevent medication errors; (2) staff their intensive care units with physician-intensivists who are more skilled at treating complex, technology dependent, very acutely ill patients; and (3) for six specific high risk conditions, perform a sufficiently high volume with adequate outcomes to enhance the safety in treating those conditions.<sup>1</sup> Those three "leaps" were announced by Leapfrog in 2001 to spur faster improvement in patient safety than history tells us would be likely without the pressure of concerted buying power; and they were included in the 30 Safe Practices measures endorsed by NQF. Leapfrog has now added the 27 remaining NQF measures as a fourth leap. These measures are as distinct from the Joint Commission's National Patient Safety Goals of 2004 for hospitals, its accreditation standards, which are aimed directly at patient safety issues, and its sentinel event policy requiring hospitals to report specific types of problems they encounter in delivering care.<sup>2</sup>

Separate from its patient safety measures,<sup>3</sup> NQF has also published a consensus report<sup>4</sup> identifying what are being referred to as "never events," serious adverse events that are

---

**[Section 4:14]**

<sup>1</sup>See <http://www.leapfroggroup.com> and Gosfield, "Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 133-38 (A. Gosfield, ed. 2004).

<sup>2</sup><http://www.jcaho.org/general+public/patient+safety/#Patient%20safety-related%20standards>.

<sup>3</sup>NQF, "Safe Practices for Better Health Care" (May 2003).

<sup>4</sup>NQF, "Serious Reportable Events in Healthcare: A Consensus Report" (2002).

avoidable failures of process that harm patients and never should occur. These include such issues as operating on the wrong patient or the wrong body part or leaving a foreign body in the patient. The matters tracked as “never events” are quite different from the quality improvement measures that are the bulk of the quality performance measurement dictionary.

**§ 4:15 Who is measuring what?—Other Measurers and Measures—IHI System Level Measures**

Yet a further take on performance measurement lies in system level measures that focus on a health care system’s overall performance on core dimensions of quality. Publicized and promoted by the Institute for Healthcare Improvement (IHI) in Boston, the theory behind this approach is that health system leaders and the public need fewer, higher level measures as opposed to a large set of highly specific measures that reflect the performance of discrete highly selected aspects of a large health care system.<sup>1</sup> Their initial set of ten measures is focused around the STEEEP aims and seeks to draw on large administrative data bases, where possible, to expedite and lower the data collection burden.

Unlike even the crosscutting measures of the already cited measurers, here the expectation is that by trending data on high-level concerns, system leaders will be directed to focus organizational energies on improvement in a very different way. Their ten measures address: (1) safety—adverse drug events per 1,000 doses; work days lost per 100 employees per year; (2) effectiveness—hospital specific mortality rate; functional outcomes; (3) patient-centeredness—patient satisfaction; percent patients dying in hospital within the region; (4) timeliness—days to third next available appointment; and (5) efficiency—health care costs per capita for the region; hospital specific standardized reimbursement (*i.e.*, Medicare allowed payments per hospital discharge). The implications of these measures go beyond a single disease, or department

---

**[Section 4:15]**

<sup>1</sup>Nelson, Nolan, Nolan, Long and Jarman, “IHI’s Health System Measures Kit:Version 1.0,” <http://www.ihl.org/IHI/Topics/LeadingSystemImprovement/Leadership/Tools/HealthSystemMeasuresKitIHITool.htm>.

within a hospital. In looking at these broader aspects of quality, one might well uncover the need to focus on the process of care for a specific disease or specific aspect of safety, staffing or some other component of quality performance, but the broader measures do not presuppose either the prioritization of the problems to address within each measured factor nor the locus of attention to improve the care analyzed by the measure.

#### **§ 4:16 Who is measuring what?—Other Measurers and Measures—And More?**

There are now so many different entities that are publishing performance measures that AHRQ has created a National Quality Measures Clearinghouse.<sup>1</sup> Taking an even broader view of the range of measures extant, this clearinghouse reports five domains of measurement: (1) access to care; (2) outcomes of care; (3) patient experience of care; (4) process; and (5) structure. Searching for “diabetes” alone produces 91 measures. The Leapfrog Group has also created a crosswalk of quality measurement sets.<sup>2</sup> In their spread sheet, they have analyzed the commonalities among measure sets from Leapfrog, Bridges to Excellence, the Integrated Health Care Association, NQF, the National Voluntary Hospital Reporting Collaborative, the CMS Premier Incentive Demonstration, the Joint Commission Oryx Core Measures and the AHRQ and patient quality indicators among others.

Although there is no question that some version of the priority conditions has been prominent in the myriad measures in play today, there are slight differences among those conditions from measures set to set, and worse yet, there is a proliferation of agencies seeking to be seen as the predominant source for measures for specific topic areas. There are both minute and greater differences among the multiplicity of measures, even as they look, at least superficially, to be approaching commonality. There is little explication, if any, as to why the variations from set to set, even as different groups endorse different constellations and combi-

---

#### **[Section 4:16]**

<sup>1</sup>[http://www.qualitymeasures.ahrq.gov/resources/measure\\_\\_use.aspx](http://www.qualitymeasures.ahrq.gov/resources/measure__use.aspx).

<sup>2</sup>*Personal Communication*, Suzanne DelBanco, Chief Executive Officer, The Leapfrog Group (Oct. 25, 2004).

nations of measures already in existence. For all of the various pronouncements from each of the bodies addressed here (and others) of a professed desire for greater unanimity and standardization, the assortment of measure sets speaks to the extremely evolutionary and transitional nature of this moment in the history of performance measurement and reporting. It is as if the prom committee has convened, but cannot agree on the theme for the event, let alone the music and decorations. Still further, the enumeration of measures activity here is not even the totality of who is involved in measures development and publication today. Nor does the abundance of published measures reflect the still considerable policy volatility regarding their uses and users.

#### **§ 4:17 Uses and users: policy controversies**

The variety of applications of performance measures to improve healthcare quality is impressive. It is in the application of measures that most controversies arise. Measures may be used for internal quality improvement to identify opportunities to change performance. They may be used in report cards to stimulate more market-based behavior from a range of actors. Payors may use the data to determine who they will allow in networks or who is eligible for increased or differential payment based on performance. Accreditation bodies incorporate data from performance measurement in their accreditation decisions. Employers and purchasers use the reports to decide which health plans, integrated delivery systems, or hospitals they will offer. Individual consumers may use measures to select their health care providers. Government agencies are using the measures in the same way the private sector is with, as noted with respect to the Medicare Modernization Act, additional programs to be developed in the near term.

All of these can have major economic significance to all of the stakeholders acting upon them or being acted upon. Obviously, there is also the implication that in measuring performance of multiple actors on the same scale there will always be a top of the class and a bottom of the class. This will create invidious comparison anxieties for those who do not score well. The point of performance measurement is either to change intrinsic behavior or to create responses because of the comparative nature of the conclusions to be drawn. Without focusing on the concerns that are generated by the

mere fact of comparative measurement, there are four fundamental areas of policy fluidity: (1) controversies over the measures themselves and construction of measurement sets; (2) the impact of the burden of measurement; (3) whether the data produced is legitimately comparable; and (4) whether the uses of measures will actually improve care.

**§ 4:18 Uses and users: policy controversies—  
Measures themselves**

The scientific content of the measures is the first point of controversy. In the distinctions between clinical practice guidelines as idealized and performance measures or indicators as consensus based, the question becomes whether the goal of consensus undermines the science. For example, in the NQF consensus development process, the products are to be voted upon, but simple majorities of those voting determine their acceptance or their rejection. Still further, all evaluations take place through the stakeholder Member Councils. However, if only two Member Councils agree on a product, it can advance for board approval. Depending on the specific perspective of the stakeholder groups on specific measures, this mechanism, while driving to an acceptable middle, can distort the results of what a more scientifically driven view might be. More to the point, participation is voluntary and self-selecting. Those who show up and speak are heard. The emphasis on articulating explicitly the scientific bases and rationales is a safeguard. So is the use of technical advisory panels and the Strategic Framework Board. The process is open and public. Still, by its very emphasis on consensus, practiced at its pure democratic extreme, such a process can be akin to voting on the sex of a cat.

The obverse of this problem is the long criticized aspect of some performance measures as being tainted with self-interest by virtue of their sponsorship and funding. David Eddy noted back in 1998<sup>1</sup> that consultants who produce measures have an incentive to create as many measures as possible. Where not for profit groups undertake measures development, there has been insufficient funding for mea-

---

**[Section 4:18]**

<sup>1</sup>Eddy, "Performance Measurement: Problems and Solutions," *Health Affairs*, 11 (July/Aug. 1998).

asures development, and sponsorship has to be carefully vetted to assure the pharmaceutical manufacturers, for example, do not seek to sponsor additional HEDIS<sup>®</sup> measures solely to further the utilization of their drug. David Lansky of FACCT takes the position that the industry itself should not be permitted to develop its own measures and only the public and regulators should.<sup>2</sup>

Some have questioned whether the specificity of and rigidity of measures ought be different depending on their ultimate applications; in other words, would measures to be used for internal quality improvement within an integrated delivery system be stated in different terms encompassing different issues than measures that would be reported to the public? The American College of Physicians states that internally focused measures should be more flexible and that external measures “demand a much higher degree of scrutiny.”<sup>3</sup> Others argue that the science and what we know and can state about it in a measure should not differ depending on its application. Formats may be refined to make data more user friendly for public consumption; and how data is presented has been a major concern for FACCT, NCQA and most other public comparison reporters; but whether the content to be reported ought to vary depending on whether it will be used publicly or privately remains undecided in the market and the policy debates.

This issue of whether measures ought to differ based on public or private use is also related to whether the measures themselves and the specifications to report them, ought to be in the public domain. Some measures are proprietary, but they will not be endorsed by NQF, for example, unless some of what makes them proprietary is relinquished.<sup>4</sup> What makes measures proprietary in the first place? The science within them is surely not proprietary, and it is their evidence base that makes them generally applicable. Should measures specifications be proprietary when the way data for a measure is collected and counted is the essence of

---

<sup>2</sup>Lansky, “Measuring What Matters to the Public,” *Health Affairs*, 40-41 (July/Aug. 1998).

<sup>3</sup>American College of Physicians, “The Use of Performance Measurements to Improve Physician Quality of Care,” A Position Paper (2004), [http://www.acponline.org/hpp/performance\\_measure.pdf](http://www.acponline.org/hpp/performance_measure.pdf).

<sup>4</sup>See § 4:8.

comparability? If the whole point is to standardize throughout the healthcare system, then aren't proprietary measures antithetical to the purpose of the broad initiatives?

#### § 4:19 Uses and users: policy controversies—Burden of measurement

Many have cited both the administrative and cost burden of reporting multiple measures to multiple sources for multiple reasons. The ACP notes that health plans and providers use multiple measures for the same phenomenon, such as HEDIS<sup>®</sup> measures, an internal measure from claims data and chart reports to measure childhood vaccination rates. Depending on the application of the measures, such as for payment, accreditation, or even participation in a government program, even though most of these efforts are characterized as voluntary, the effect of the reporting initiative can be such as to *mandate* participation in practical terms. Yet CMS, Leapfrog, the Joint Commission, and HEDIS<sup>®</sup>, while addressing the same priority conditions, are asking for slightly different constellations of reports. The data must be reported to multiple entities and in multiple formats. Intuitively, it is inevitable that this alone creates both administrative and cost burdens. The Leapfrog Group's voluntary hospital survey when printed from the web is 111 pages long!

Still further, while some data can be gleaned from claims data (*e.g.*, ordering of laboratory studies or prescribing drugs), most of the activity being measured by most of the measures will only be documented in the patient medical record. There are too few electronic medical records that create the capacity to easily report and no single data set will satisfy all sources. The burden of the technology gap has often been cited.<sup>1</sup> Yet, there is virtually no data in the published literature on the quantification of the burden.

---

#### [Section 4:19]

<sup>1</sup>Eddy, "Performance Measurement: Problems and Solutions," *Health Affairs*, 11 (July/Aug. 1998); Adams and Corrigan, eds., "The 1st Annual Crossing the Quality Chasm Summit" (National Academic Press 2004); American College of Physicians, "The Use of Performance Measurements to Improve Physician Quality of Care," A Position Paper (2004), [http://www.acponline.org/hpp/performance\\_measure.pdf](http://www.acponline.org/hpp/performance_measure.pdf); Comptroller General's Forum, "Unsustainable Trends Necessitate Comprehensive and Funda-

The burden of measurement relates as well to the maxim that “what gets measured gets done.” Health care entities that are measured will perform to meet the measure. The more that is measured the greater the work to perform to the measure. This maxim now finds evidence in a recent study that showed that in the Veteran’s Administration (VA) system, which had introduced both an electronic medical record and performance measurement, the VA patients fared considerably better (by 15 and even 20 percentage points) than a national comparable cohort group not explicitly subject to measurement.<sup>2</sup>

The corollary maxim “be careful what you measure” reflects the belief that where scarce organizational resources will be devoted to performing to the measure, performance on other unmeasured dimensions may degrade. While the maxim is often put forth by those who are measuring, there is far less evidence that the quality of other care will diminish. But the same study of the VA found that on dimensions of care in the VA that were not measured, the differences between the national unmeasured sample and VA’s performance “barely reached conventional levels” of statistical significance.

The burden of measurement means that in an era of scarce resources to devote to the very essential task of improving quality, selecting a parsimonious set of measures is important. While this principle is widely broadcast in the industry, it is being honored more in its breach in the public policy discussions. When health system leaders and managers seek to select among the myriad available measures for internal improvement, they face a daunting array of possibilities. Burden will be an important consideration; but clarity regarding the purpose of using the measure and what will be done based on the data adduced will enhance those choices. “[Measures] are the cheese, not the whole

---

mental Reforms to Control Spending and Improve Value,” GAO-04-793SP, 18 (May 2004).

<sup>2</sup>Asch, McGlynn, et al., “Comparison of Quality of Care for Patients in the Veteran Health Administration and Patients in a National Sample,” *Ann. I. Med.*, 938-45 (Dec. 21, 2004).

sandwich.”<sup>3</sup> In other words, it is a waste of time and resources to measure things upon which you will not act. Although NQF itself and the statements of CMS, the AMA, and virtually everyone who speaks on the topic is to develop uniform standardized measures, the mere existence of the AHRQ measures clearinghouse speaks to the Tower of Babel effect on those who are the subject of measures. Beyond internal measurement, there is another policy problem in the comparability of what is being reported.

#### § 4:20 Uses and users: policy controversies—Are they comparable?

The first and now classic comparability issue often raised in performance measurement is whether the measures are appropriately refined to reflect adequate risk adjustment. “But our patients are sicker” has long been the response of those who do not meet expectations when they are reported upon. The ACP has called for demonstration projects reporting on physicians but only when the data are “fully adjusted for case-mix composition (including factors of sample size, age/sex distribution, and severity of illness; number of comorbid conditions; and other features of a physician’s practice and patient population that may influence results.”<sup>1</sup> Whether the data is comparable turns as well on the accuracy of what is reported:

As for inaccuracy, it can creep into performance measurement through every pore. Some of the most obvious sources are insufficient sample sizes, inaccuracies in the data sets, and the presence of confounding factors that are either understood but not adjusted for or not understood at all.<sup>2</sup>

Although measures may be very similar in their scientific content, the sources of data from which the measures are

---

<sup>3</sup>Reinertsen and Schellekens, “Ten Powerful Ideas for Clinical Quality Improvement,” Health Administration Press (forthcoming 2005).

#### [Section 4:20]

<sup>1</sup>American College of Physicians, “The Use of Performance Measurements to Improve Physician Quality of Care,” A Position Paper, 9 (2004), [http://www.acponline.org/hpp/performance\\_measure.pdf](http://www.acponline.org/hpp/performance_measure.pdf).

<sup>2</sup>Eddy, “Performance Measurement: Problems and Solutions,” *Health Affairs*, 11, 16 (July/Aug. 1998).

obtained can significantly skew the outputs.<sup>3</sup> It is for this reason that NCQA's HEDIS<sup>®</sup> program moved from a self-reported phenomenon to one where the specifications for measures are very detailed and there is a coterie of approved validators who will certify the data as having been created in the appropriately comparable manner.<sup>4</sup> Without standardized specifications for measures, standardized mechanisms to report (as opposed to the resulting standardization in the reports) all of these many measures and the reports they will generate cannot be assumed to be comparable even when reporting on the same dimensions of quality.

Another problem is that some of the high profile initiatives still entail predominantly self-reported data. For example, the Leapfrog survey is essentially self-reported. The Bridges to Excellence program requires physicians to glean their own data from their own medical records and has a very low validation percentage.<sup>5</sup> How do we know we are getting a true picture of what has been done, if the targets self-report? Can self reported data be comparable to validated data?

#### § 4:21 Uses and users: policy controversies—Quality improvement

All of these concerns, though, pale against the overarching issue for all performance measurement initiatives: whether the application of measures will actually improve the delivery of care. It may be that because of the demand for consensus (and antipathy for regulation)<sup>1</sup> in today's world we are really only addressing very low hanging fruit on the quality tree. To the extent that the targets of measurement

---

<sup>3</sup>Byrnes, "Data Collection" in Nash, et al., *The Healthcare Quality Book*, 117-43 (Health Administration Press 2004).

<sup>4</sup>See, e.g., "NCQA's HEDIS Compliance Audit Program and Software Certification Program," <http://www.ncqa.org/Programs/HEDIS/Hedis%20compliance%20audit%20program%20and%20sc.htm>.

<sup>5</sup>See Gosfield, "Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 103-82 (A. Gosfield, ed. 2004).

#### [Section 4:21]

<sup>1</sup>See Schlesinger, "On Government's Role in The Crossing of Chasms," 29 *J. Health Pol. Pol'y and L.* 1 (Feb. 2004) (lambasting as "comical (albeit tragicomic)" the Crossing The Chasm report for shying away from recommending any regulatory role for government in transforming the quality of health care for the 21st century).

change their behavior on those dimensions of care, and the measures are appropriately grounded in science, then this should actually improve the quality of care delivered.

But there is evidence that only those who are performing at the lowest tier of the class actually do change their behavior in response to public reporting programs and those who are in the middle or even above the middle, are not so motivated.<sup>2</sup> In New York when data was published regarding hospital and surgeon performance of cardiac surgery, there was not one change in market share among those reported on, even for those at the utter bottom of the class. Payors did not deselect them,<sup>3</sup> nor did patients choose other institutions and physicians.<sup>4</sup> Still, the mere fact of disseminating the information publicly, clearly sparks some response, since other studies have shown that in states where there have been public CABG report cards profiling the providers who perform these services, care does improve.<sup>5</sup>

Some would take the position that if we are really going to improve care, public reporting is essential because the accountability it generates by itself will per force be more effective than professionalism alone which cannot be expected to change behavior in significant ways or it would have already.<sup>6</sup> Whether the public disclosure of quality data improves care, and whether the measures reported publicly

---

<sup>2</sup>Chassin, Achieving and Sustaining Improved Quality: Lessons From New York State and Cardiac Surgery, *Health Affairs*, 40-51 (July/Aug. 2002).

<sup>3</sup>Chassin, "Achieving and Sustaining Improved Quality: Lessons From New York State and Cardiac Surgery," *Health Affairs*, 40-51 (July/Aug. 2002). This finding contradicts earlier data that hospitals with better outcomes had experienced higher rates of growth in market share. See Mukamel and Mushlin, "Quality of Care Information Makes a Difference," 36 *Medical Care* 945 (1998).

<sup>4</sup>But there has long been evidence that, least to date, consumers have not changed their behavior based on public reporting either. See Gosfield, "Health Care Report Cards: Quality in the Public's Cross Hairs," *Health Law Handbook*, 501-42 (A. Gosfield, ed. 2000).

<sup>5</sup>Hannan, et al., "Provider Profiling and Quality Improvement Efforts in Coronary Artery Bypass Graft Surgery," 41 *Medical Care* 1164-72 (2003).

<sup>6</sup>Marren, Feazell and Paddock, "The Hospital Board at Risk and the Need to Restructure the Relationship with the Medical Staff: Bylaws, Peer Review and Related Solutions," 12 *Annals Health L.* 179 (2003). This argument is made in the context of why the hospital board should control

should differ from those used internally, is still a contentious issue in the performance measurement debates. In examining the available data on point, in the context of what the United Kingdom might learn from American efforts, the landscape was found to be “an evidence sparse policy zone.”<sup>7</sup> The reviewers there found that “surprisingly little research has been conducted into the utility, acceptability, and impact of public disclosure in the US.” Since that review five years ago, given the new applications of performance measures reporting, we can expect more evidence soon, although it would be difficult to predict in which direction it will point. Even in New York where the poor performers had dramatic improvement, and, by comparison with other states reporting similarly, demonstrated far more relative improvement, the reasons are believed not to relate to the public nature of the reporting itself but to six elements brought to bear together at the instigation of the State Department of Health: (1) required reporting from all hospitals; (2) regular audits to verify data quality; (3) analysis and public reporting by a neutral, respected third party; (4) close oversight by an advisory group of recognized clinical leaders; (5) a commitment to studying and publishing reports on the impact of the system; and (6) continuous pressure on poor and mediocre performers to improve.<sup>8</sup>

Given the very real questions, though, as to the effect of publicly reporting performance measures, it is also worth looking briefly at the potential effects on quality where measures are intended purely for internal uses. There is some data that profiling of physicians within hospitals and health care institutions to direct attention to outliers can

---

quality more directly than has traditionally been the case in the self-governing medical staff relationship to the Board. I don't agree with the specific solutions proposed in this article, but the principle that self-regulation has not gotten us where we need to be is fairly unassailable.

<sup>7</sup>Marshall, et al., “Public Disclosure of Performance Data: Learning from the US Experience,” *Quality & Safety in Health Care*, 53-57 (2000), <http://qhc.bmjournals.com/cgi/content/full/9/1/53>.

<sup>8</sup>Chassin, “Achieving and Sustaining Improved Quality: Lessons From New York State and Cardiac Surgery,” *Health Affairs*, 40, 49-50 (July/Aug. 2002).

produce improvement.<sup>9</sup> Some prefer that comparative data be kept within the industry for purposes of benchmarking, a practice which compares the results for one organization or institution to others who have achieved excellence or are top performers.<sup>10</sup> However, to the extent that Americans are getting only 55% of the care that science says they should<sup>11</sup>, the overall class is generally not doing very well in terms of providing evidence based medicine. Benchmarking within an overall poorly performing class has been described as measuring against the “cream of the crap,”<sup>12</sup> hardly what the ultimate goal should be.

Are there other pathways? Some would argue that aiming at system level measures is particularly important for deep and broad system-level improvement.

When you try to change performance based on system-level measures, such as hospital mortality rates, the improvement agenda moves far beyond evidence-based care for a few diseases, and into the culture of the institution, the leadership system, the infrastructure of human resources, finance and information technologies; in other words, you wind up having to transform a whole organization. These measures therefore become an appropriate means of holding the senior leadership of an organization to account.<sup>13</sup>

But this presupposes a real commitment to making change based on the difficult information that such in depth analysis can provide. The real point is that it is unlikely that the mere fact of performance measurement in any setting will by itself produce improved quality. It is in the action taken upon the reported data that measures will have their impact, including legal impacts.

---

<sup>9</sup>Nash and Evans “Physician and Provider Profiling” in Nash, et al., *The Healthcare Quality Book*, 167-81 (Health Administration Press 2004).

<sup>10</sup>See, e.g., Lee, et al., “Statistical Tools for Quality Improvement” in Nash, et al., *The Healthcare Quality Book*, 164-65 (Health Administration Press 2004).

<sup>11</sup>McGlynn, et al., “The Quality of Health Care Delivered to Adults in the United States,” 548 NEJM 2635 (Jan. 16, 2003).

<sup>12</sup>Private communication James Buckman, Executive Director, Juran Quality Institute, Carlson School of Management, University of Minnesota, to James L. Reinertsen, M.D. (Aug. 2003).

<sup>13</sup>Reinertsen, *Personal Communication* (Dec. 29, 2004).

**§ 4:22 Legal Issues**

There is a dearth of regulation dealing with performance measurement; although the federal government is increasingly involved in performance measurement, and state agencies have been publishing report cards for years. The legal issues in performance measurement will lie primarily in: (1) the legal status of measures and measure sets defined by different sources; (2) legal issues that arise in applying measures; and (3) liabilities in measurement initiatives themselves.

**§ 4:23 Legal issues—Constructing measures and sets**

The first obvious legal issue for those who would create measures, construct measurement sets, and publish data based upon those measures is whether there is a standard of care that has developed with respect to these activities. One might expect potential legal vulnerability to a charge of negligent measures construction where those who are reported upon are damaged by the impact of the reports. When the IOM published its eight attributes of good CPGs, they very shortly became generally regarded as establishing the standard of care for CPG construction. The attributes of good performance measures have also been stated by the IOM even though they were designated in 1992 as attributes of “medical review criteria.” Those attributes were refined in part in the development of HEDIS<sup>®</sup>. The President’s Commission also spoke to factors that would be important in selecting measures.<sup>1</sup> There has been no further articulation since of better, decidedly different or more specific attributes of good measures. Whether measures which cannot claim to manifest the attributes stated by the Commission will be open to challenge as negligently constructed is an untested issue.

One might have expected to find a more recent iteration of what would qualify as a standard for good performance measures in the work of the NQF. But the NQF does not prejudge what it will endorse. It relies on the consensus process to specify what will be endorsed. On that basis, the

---

**[Section 4:23]**

<sup>1</sup>See § 4:2.

outputs of NQF may well differ from what the IOM or President's Commission attributes would produce on the same topics. Whether any measure proposer will take legal action if their measures are not endorsed but are purported to be 'better' is another potential legal quagmire. There is an additional question as to whether the consensus process alone, gives the NQF endorsed measures different legal status.

The requirement in the Technology Transfer Act that federal agencies turn to voluntary consensus standards organizations for the bulk of their standards reflects a Congressional bias toward private enterprise, even where one of the ultimate outlets for applying standards will be through government programs. How much of a free market in measures we need is an unstated aspect of NQF's work, especially given the driving policy goal of standardized measures and specifications. To select measures to endorse, though, implies choice and therefore some competition among measures and measurement sets. NQF has explicitly stated it is seeking to preempt the field with its endorsed measures. Since NQF does not create measures but endorses those proposed to it, the paradoxical effect may be to set up NQF as the ultimate arbiter of what will prevail in the market. This is a decidedly different legal status from what any professional organization, health care system, payor, research agency or consulting organization might develop. While the actual legal status of the measures themselves is speculative, it is less difficult to imagine the legal issues arising over the applications of measures and their impacts.

#### **§ 4:24 Legal issues—Legal challenges to measures applications**

Measures used in internal quality improvement may establish the basis for a hospital medical staff to limit, restrict, or terminate a physician's clinical privileges. In fact, to use explicit performance measurement in this way would be an improvement over the usual idiosyncratic approach to these

issues in most hospitals.<sup>1</sup> If that were to happen, whether the standards for taking such action were made known and applied fairly would be an issue raised by the aggrieved physician. Therefore, measures selection, the standards used to apply them, and how they are applied to the relevant class of physicians are likely battlegrounds, as hospital medical staffs move more to these kinds of evaluative techniques. Where the QIOs, NCQA and the Joint Commission take negative actions based in part on the data in their performance measurement, it can be expected that the same types of issues with respect to the quality and reliability of the data and the conclusions drawn based on the data will likely be raised; but it is far less likely that the measures themselves will be challenged.

Similarly, to the extent that payors use performance measurement to determine who will continue in their networks, terminated providers may look at the legitimacy of the performance measurement program and whether utilizing these mechanisms has been accounted for in the provider contract. Those contracts typically require provider compliance with quality improvement programs and cooperation with the plan's HEDIS<sup>®</sup> needs; but they rarely disclose in their contracts or manuals their internal measures for network inclusion or exclusion, nor for quality review. How plans use performance measurement in the contract termination process may be an issue going forward, although most plans finesse any judgment of quality by simply terminating without cause. The plan sponsored P4P programs are a significant departure from these traditional approaches.<sup>2</sup>

In P4P programs the plans have provided extremely explicit statements of the performance measures that provide the basis for the judgment that additional payment should be made. Most of the measures come from HEDIS<sup>®</sup>, but some are additional. Still, however, legal challenges to the payment determinations (whether payment should be made at

---

**[Section 4:24]**

<sup>1</sup>See Gosfield, "Whither Medical Staffs: Rethinking the Role of the Medical Staff in the New Quality Era," *Health Law Handbook*, 141-218 (A. Gosfield, ed. 2003).

<sup>2</sup>For more information about P4P programs, see Gosfield, "Contracting for Provider Quality: Then, Now and P4P," *Health Law Handbook*, 103-82 (A. Gosfield, ed. 2004).

all and if so, how much) likely will not arise since: (1) many of these programs, including Bridges To Excellence, are not grounded in contract at all; (2) those that are grounded in contract often deny to payees the right to appeal the amount of any payment or the refusal of payment; and (3) many of these programs tier the participants so that there is no way of knowing whether any payment is owed unless the plan is willing to reveal the ranking of the participants.

Other developing applications for performance measures include denied payment for basic services by MCOs as in the new program of HealthPartners in Minnesota which is refusing to pay hospitals for any of the 27 “serious reportable events” known as “never events” as designated by NQF.<sup>3</sup> Even in light of a new state law mandating reporting of these events to a state agency, HealthPartners is asking hospitals to report these events to the health plan directly in addition to government notification. These denials will likely be appealed, if at all, through standard plan payment appeals processes.

The contractual context for performance measures is where most, if any, of the legal challenges to performance measurement will likely lodge, and they will be about what happens based on the measurement and not the measurement itself. The measurement process or failure of it raises still different legal concerns based in tort.

#### § 4:25 Legal issues—Liabilities from measurement

The fact of performance measurement in the industry is likely to create new forms of liability in as yet unrealized ways. Consider the fact that the measures are being used to differentiate providers and, of necessity, will state comparative performance. Whether the content of the measures states reasonable, expected performance and therefore a legal standard of care is a real issue, since much of what is being measured is still aspirational. The point of the Leapfrog Group standards was to push providers to adopt programs and systems that, by definition, were not yet standard behavior. Some have speculated that it is entirely possible that the Leapfrog Group standards will make their way

---

<sup>3</sup>Kazel, “Insurer Won’t Pay Hospital for ‘Never Events,’ ” *American Medical News*, 20 (Nov. 8, 2004).

into plaintiffs' cases against hospitals that do not embrace the four leaps.<sup>1</sup> While big tort awards on such bases might motivate hospitals to adopt the activities measured as a risk management technique, the fear of lawsuits is hardly a sufficient motivation to really propel the quality changes the standards are intended to generate. If risk anxiety were a sufficient motivation, the patient safety measures would never have been necessary, since the risks they address have long been actionable. But those standards implicate more traditional tort liability issues since they are about patient safety. On other dimensions of quality, the implications are different, and it is noteworthy that some in quality policy world have taken a harder line with regard to the legal import of quality measures.

The godfather of documenting unexplained variation, John Wennberg, has written a law review piece<sup>2</sup> (with an attorney) in which he argues that the quality of the evidence upon which measurement is based should drive to different liability outcomes. They differentiate three different categories of care measures: (1) effective care; (2) preference-sensitive care; and (3) supply side care. Effective care is services whose use is supported by well-articulated medical theories and strong evidence of efficacy. With respect to this type of care the answer to the question "which rate is right?" would be "100%." Here they include HEDIS<sup>®</sup> measures. Preference-sensitive care is where the evidence would offer a choice between at least two treatments with different risks and benefits, as in the choice between a lumpectomy and a radical mastectomy. Supply-sensitive care refers to services that are generally provided in the absence of medical evidence and specific clinical theories about the benefit gained relative to the frequency. In a sense, they are addressing in effective care rates problems of underuse; in preference sensitive care, patient centeredness; and in supply-sensitive care, the problems of overuse.

---

**[Section 4:25]**

<sup>1</sup>Mello, Studdert and Brennan, "The Leapfrog Standards: Ready to Jump from Marketplace to Courtroom?," *Health Affairs*, 45-59 (March/April 2003).

<sup>2</sup>Wennberg and Peters, "Unwarranted Variations in the Quality of Health Care: Can the Law Help Medicine Provide a Remedy/Remedies?," 37 *Wake Forest L. Rev.* 925 (2002).

From a liability perspective, they argue that for effective care,

when there is strong evidence and wide professional and societal consensus about appropriateness, the legal standard of care should be that virtually all of those in need of such care should receive it. Courts should treat departure from that consensus as negligence, *even if most physicians depart from the proven approach*.<sup>3</sup>

On the issues of preference sensitive care, courts should move, they agree, from the traditional doctrine of informed consent to a doctrine of informed patient choice. The former puts the emphasis on what the physician considers material and the latter on what the patient considers material. On the supply-sensitive care, plaintiffs would be arguing that the care is of a type regularly over prescribed and it was not medically necessary in this specific case with reference to aggregate data showing the physician should have used a more conservative standard of care. But lower utilizing physicians might, they speculate, be able to justify not providing the unnecessary care using the same data.

These types of arguments point to the considerable anxiety providers have with respect to their enhanced risk of lawsuit if negative, comparative data about them becomes known. Very recently the Florida Hospital Association sought an injunction to prevent disclosure of adverse hospital incidents after a new constitutional amendment making such data available to the public.<sup>4</sup> The case was dismissed for failure to state a justiciable controversy. In contrast, provider anxiety as a barrier to improvement has been confronted in a new law in the State of Washington that protects small physician groups that establish Coordinated Quality Improvement Programs to review negative outcomes, injuries, infective controls and quality improvement strategies. Their data can be protected from discovery and shared with other

---

<sup>3</sup>Wennberg and Peters, Unwarranted Variations in the Quality of Health Care: Can the Law Help Medicine Provide a Remedy/Remedies?, 37 Wake Forest L. Rev. 925, 933 (2002) (emphasis added).

<sup>4</sup>Florida Hosp. Ass'n v. Florida Agency for Health Care Admin., No. CA 002670 (Fla. Cir. 12-7-04); see "Florida Court Rejects Hospitals' Bid to Block Access to Medical Error Data," 13 BNA, HLR 1788 (12-16-04).

organizations.<sup>5</sup> The public versus private reporting, internal improvement versus external accountability uses debates in performance measurement are unresolved. Risk of disclosure of data demonstrating poor performance is one problem. Almost the obverse risk exists where data is intended to be used only internally, but the data collectors do not review and do not act upon data they have.

That hospital boards should be collecting important data about quality of care and exercising the will to act upon it is part of their fiduciary responsibilities in stewarding the assets of the businesses they run.<sup>6</sup> To the extent they are reporting to their boards, however, most hospital administrators are either overwhelming them with too much data that is not significant or “the Board receives scant and often meaningless data with little or no insight into the clinical quality challenges and opportunities of the organization.”<sup>7</sup> Pitfalls lurk where performance data languishes unaddressed.

In the case of *Cronic v. Doud*,<sup>8</sup> the appeals court overturned the lower court’s grant of summary judgment to the hospital in a malpractice case. Critical to the reversal was the court’s focus on the allegations that the hospital had been maintaining data about utilization and should have seen that a high volume of surgeries demanded, if not just attention, potentially even termination of the physician co-defendant’s surgical privileges. The court reiterated the duty of the hospital to know the qualifications and the standards of the performance of its physicians. But to breach the duty, the hospital must be put on notice of the problem. The allegations regard-

---

<sup>5</sup>Robeznieks, “Law Protects Physicians with Patient Safety Plans,” AMNews, 13 (Nov. 8, 2004).

<sup>6</sup>Reinertsen, “Understanding and Improving Clinical Quality: The Role of Trustees,” *Trustee* (Aug. 2003). A 2003 change in the Medicare Hospital Conditions of Participation now requires hospitals to develop, implement and maintain a data driven quality assessment and performance improvement program. 42 C.F.R. § 482.21. The program must include quality indicator data and the Board is held accountable for the operation of the program. 42 C.F.R. § 482.21(e).

<sup>7</sup>Marren, Fezell and Paddock, “The Hospital Board at Risk and the Need to Restructure the Relationship with the Medical Staff: Bylaws, Peer Review and Related Solutions,” 12 *Annals Health L.* 179, 219 (2003).

<sup>8</sup>*Cronic v. Doud*, 168 Ill. App. 3d 665, 119 Ill. Dec. 708, 523 N.E.2d 176 (4th Dist. 1988).

ing the data demonstrating an increase in surgery created a sufficient issue of fact to permit the case to proceed to trial.

Similarly, as fraud and abuse liabilities attach to overuse problems, the failure to act on performance data could lead to false claims liability. In the Tenet hospital chain's problems generated by purportedly unnecessary cardiac cathardizations and bypass surgeries in Redding, California, the overuse of procedures was at issue. Still further, in the United Memorial Hospital settlement<sup>9</sup> overuse of services was also at issue.

Still another potential predicament may arise from faithful reliance on the priority conditions as setting forth the standard of care with respect to the scope of performance measurement activities in which organizations ought be engaged. The publication of priority conditions now endorsed by NQF was intended to spur concerted action across the country on specific dimensions of healthcare quality. Most of the priority conditions are disease focused but several of the conditions identified by the IOM and endorsed by NQF are more crosscutting in their impact. While the point of articulating these conditions was to focus scarce resources to produce important changes throughout the system, reference to the priority conditions as driving what organizations do in terms of internal measurement will not offer real legal solace.

There is a powerful argument that focusing on a different approach, particularly within hospitals and systems would be more important, at least in legal terms. Some would argue that paying attention by measurement to the fundamental issue of hospital mortality rates and their causes is far more important.<sup>10</sup> There is no question that measures focused on these targets would drive to a very different analysis from what would emerge from tracking data on the priority conditions, which are, purposefully focused on chronic care. There is increasing evidence as to the value of system level data,

---

<sup>9</sup>Plea Agreement, *United States v. United Mem'l Hosp.*, No 1:01-CR-238 (WD Mich 1-8-03). *See also* Sanderson, "Hospital Chief May Soon be Facing Trial," *Traverse City Record Eagle* (Jan 11, 2003), available at <http://www.record-eagle.com/2003/jan/11cheb.htm> (accessed Mar. 25, 2004).

<sup>10</sup>Reinertsen, "Understanding and Improving Clinical Quality: The Role of Trustees," *Trustee* (Aug. 2003).

and, more to the point, on some very simple control features that can reduce mortality by very significant percentages,<sup>11</sup> for issues such as deaths of patients on ventilators, or prevention of surgical wound infections. Performance measurement which focuses on the dimensions of care that make a difference between whether the patient is dead or alive on discharge may rise to a level of creating an almost product liability type of exposure for hospitals who fail to investigate their mortality rates and implement known corrective actions. Certainly a hospital's failure to track, review and correct for "never events" would be a potent weapon for a plaintiff seeking to show negligence on the hospital's part.

Some managers, including physicians, faced with these dilemmas, will choose to shy away from performance measurement because they are risk averse. These are the managers who approach the implementation of explicit evidence-based medicine standards from the position that if they articulate what should be done and then fail to act accordingly they are increasing their legal exposure. This is akin to putting your hands in front of your face and believing no one can see you. Articulating the proper care as an explicit management approach does not increase risk one iota. The articulation of proper care does not create a standard. The standard exists whether the management strategy to improve is stated or not. Those who take this position misperceive the fact that the point of performance measurement is to improve quality, which will improve their risk exposure. To the extent that performance measurement provides a basis on which to improve care, if acted upon it will decrease the events that lead to suits, which is the whole point anyway.

#### § 4:26 Conclusion

We need to improve the quality of American health care. There is no question that performance measurement is on the rise as a critical health care quality improvement technique. If "what gets measured, gets done" and "you cannot improve what you don't measure," then performance

---

<sup>11</sup>Berwick and Joshi, "Health Care Quality and the Patient," in Nash, et al., *The Healthcare Quality Book*, 3-24 (Health Administration Press 2004), and Reinertsen, "Understanding and Improving Clinical Quality: The Role of Trustees," *Trustee* (Aug. 2003).

measurement is here to stay. But for all the aphoristic beliefs and good motivations regarding implementation of performance measurement, much is unsettled.

Will focus on priority conditions bring dramatic care improvement to patients with chronic conditions whom the American health care system has slighted until now? Will the increased public policy emphasis on transparency of what is measured finally create a quality improvement imperative throughout the system, even if it does not create a ‘market’ in quality? Will system level measures improve care more quickly and deeply in health care organizations that use them than those who deploy the many other discrete measures of care?

Even in the face of these unknowns, the marketplace is by no means letting the perfect be the enemy of what we have now. There is competition among measures and measurers. There is an enormous amount of noise among the players clamoring for primacy in standardization. Instead of agreement and consensus we have what appears to be an ever expanding body of “hundreds of process measures, guidelines, safety indicators, and service quality measures that make up the measurement web entangling any given hospital, multispecialty group practice, or other care delivery “system.”<sup>1</sup> The plethora of reporting demands in a policy volatile context has to be increasing the burdens on those subject to measurement. This cannot help the quality improvement cause in the short term. In the longer term, there is a real question as to whether letting a performance measurement free market reign until things settle out is a good expenditure of scarce health care resources and energies which should be devoted to driving unceasingly to deliver care which meets the STEEEP aims. Yet there is no obvious locus for, nor likely benefit from, a regulatory solution. It may be that the increasing emphasis on measurement in Medicare’s programs may drive the measures “market” more quickly than it would coalesce, left to its own behavior, but this is speculative at best.

At the end of the day, probably the best that can be said about the state of the art of performance measurement is

---

[Section 4:26]

<sup>1</sup>Reinertsen, *Personal Communication* (Dec. 29, 2004).

that it is evolving. The music is playing but it is hardly decided as to whether the preferred style of dance will be ballroom, Latin, hip-hop, country or the event will devolve into nothing better than a mosh pit. More instruments keep joining the band and more dancers are crowding the hall. The transitional moment in which we find ourselves needs some stimulant, though, to get us beyond the current tumult or we will lose significant opportunities to make a difference in the health care Americans get by over-burdening the environment with too much and too many measures. It may be that as more quality performance information becomes publicly available—whether measured comparably or not, on the same conditions or not—about how the health care system is lagging behind optimal levels, a public clamor for improvement will force the issue. We can hope.