

# Performance Measures for Health Care Systems

David R. Nerenz, Ph.D.  
Michigan State University

Nancy Neil, Ph.D.  
Virginia Mason Medical Center

Commissioned Paper for the  
Center for Health Management Research

May 1, 2001

# Performance Measures for Health Care Systems

## Objectives

In this paper we will briefly summarize the history of performance measures for hospitals, health plans, and health care systems. We will describe some of the key design considerations that go along with successful performance measurement systems, and highlight some specific examples of particularly successful and/or innovative measurement systems. We hope that the information will serve as a template for health care organizations that wish to enhance their own performance measurement programs.

In discussing the history and current status of performance measurement systems, we will try to keep several key questions in mind, since they underlie virtually all other considerations about performance measures:

- ♦ *What is the entity being measured?* Which individuals and which organizational units will see their behavior reflected in any particular performance measure?
- ♦ *Who is using the information?* Is the information to be used primarily by senior managers for oversight, compensation decisions, and strategic planning? Is the information to be used by “line workers” for quality improvement? Is the information to be used by external stakeholders for contracting purposes?
- ♦ *What core organizational processes or skills are the measures designed to reflect?* Are they primarily about clinical quality of care, or management efficiency, or profitability, or something else?

## Brief History of Performance Measures in Health Care

### Private Crusader

The history of performance measurement in health care goes at least as far back as Florence Nightingale in the middle of the 19<sup>th</sup> century.<sup>1</sup> Ms. Nightingale was concerned about sanitary conditions in hospitals, in both military hospitals in the Crimean War and in London. She developed an elaborate data collection and statistical analysis system focusing primarily on *in-hospital mortality*. State-of-the-art (at least for the time) graphical presentations were used to highlight key findings for audiences that would not have been comfortable with large tables of numbers or dense text. This system allowed comparisons from hospital to hospital, from unit to unit within hospitals, and within the same hospitals over time. Having an explicit, objective measurement system allowed Ms. Nightingale and others to make significant breakthroughs in the understanding of the relationships between sanitary conditions and hospital morbidity and mortality.<sup>2</sup>

### Public Crusade

Dr. Amory Codman, a Boston surgeon, extended these ideas in a crusade for public reporting of hospital mortality data in the Boston area for a period of several years between 1914 and 1920. He developed a system of categorizing the presenting complaint and type of surgery performed for each of his patients, then tracking their course over time to determine outcomes as defined by mortality and morbidity. When his system was not accepted by colleagues in the hospital where he practiced, he established his own hospital based on the concept of “*end results*.”

Codman’s enterprise was modestly successful for a while, but closed eventually and the concept of widespread public reporting of hospital outcomes didn’t re-emerge until 1986 when HCFA began public reporting of hospital mortality rates.<sup>3</sup>

### Outcomes Management

In 1988, Dr. Paul Ellwood issued a call for a national program of “*Outcomes Management*.”<sup>4</sup>

His vision was not specific to hospitals or any other entity, but was based on the core concept of health care providers being accountable for patient outcomes. The term outcomes included not only mortality and objectively defined morbidity, but also “softer” concepts like patient-reported pain, functional status, and quality of life. Dr. Ellwood envisioned a set of large national data bases that would include commonly defined clinical, demographic, treatment, and outcome variables for major clinical conditions. Individual clinicians would be able to access the data base and obtain information on which treatments had produced the best outcomes for patients like the one they were treating. The data base would serve as a form of “collective clinical experience” for thousands of individual practitioners.

The American Medical Group Association has sponsored a number of condition-specific projects (e.g., diabetes, asthma, joint replacement surgery, cataract surgery) based on Dr. Ellwood’s concepts,<sup>5</sup> but the vision has never taken hold on a wide scale. There are a number of collaborative surgical outcome projects (e.g., Northern New England Cardiac Surgery Group, MODEMS) and medical treatment outcome registries (National Registry of Myocardial Infarction) that are very similar to Dr. Ellwood’s Outcomes Management concept. These systems, though, are primarily designed to assess performance of a single surgeon, surgical technique, or a medical treatment rather than to assess the performance of an organization.

### Performance Measures for Managed Care

Shortly after the publication of Dr. Ellwood’s landmark article a group of major national health care purchasers met to develop a set of *performance measures for managed care plans*. The measure set came to be known as the Health plan Employer Data and Information Set (HEDIS), and was adopted by the National Committee for Quality Assurance (NCQA) as the prototype performance measure set for HMOs. HEDIS went through an early

development and refinement phase, and then was tested in a national demonstration project in 1994. It became part of NCQA's health plan accreditation process shortly thereafter, and is currently a clear national standard for performance reporting for managed care plans. Participation in HEDIS is now required for plans seeking NCQA accreditation, and most plans allow NCQA to report their annual HEDIS data publicly.<sup>6</sup> However, HEDIS measures focus mainly on preventive and primary care services for defined populations of health plan enrollees; few, if any, HEDIS measures are interpretable as measures of hospital performance or of multi-hospital system performance.<sup>7</sup>

### Processes and Outcomes

In the same time period (late 1980s, early 1990s), the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) developed its own performance measurement initiative, the "Agenda for Change." An extensive set of *measures of clinical processes and patient outcomes* was pilot tested in volunteer hospitals for several years, with a major effort planned for 1994 and 1995. When JCAHO proposed to make many of these measures mandatory for hospitals seeking JCAHO accreditation, though, there was a major backlash from hospitals objecting to both the burden of data collection and the potential for inaccurate and misleading reports. Hospitals which had already made commitments to other kinds of clinical measurement systems from private vendors objected to the need to do major overhauls of existing data collection and reporting systems.

These objections led to a significant crisis at JCAHO in 1994 and 1995, and JCAHO had to abandon the plan for a standard national performance reporting system for hospitals. Instead, JCAHO developed the ORYX system.<sup>8</sup> ORYX is not a single measurement system, but rather an approved list of measures and measurement systems. Hospitals are required to participate in a minimum number of measures from one or more of those on the approved list, and to have the "performance measurement vendor" send information to JCAHO on a regular basis. The approved measurement

systems must have enough participating hospitals to allow comparisons to be made from any one hospital to a group of hospitals using the same system.

JCAHO has also developed accreditation programs and recommended measure sets for "health care networks." This can include integrated health care systems, managed care plans, long-term care networks, or other types of multi-entity systems.<sup>9</sup>

The mid-1990s also saw the development of the Foundation for Accountability (FACCT), a *purchaser-driven initiative to focus performance measures on patient outcomes* rather than on other aspects of performance. FACCT has developed standard measure sets for a number of clinical conditions (diabetes, depression, breast cancer) which include a combination of clinical, demographic, and patient-reported outcome variables. A number of demonstration projects have been conducted to test the utility of these measures, but the measure sets have not been adopted widely as a national standard for assessing patient outcomes.<sup>10</sup>

### Paying for Performance

Coalitions of public and private purchasers have maintained a steady interest in performance measurement programs that would allow them to compare hospitals, health plans, and health care systems. In some cases these systems are intended to directly compare competing providers in a single market area for purposes of competitive contracting; in other cases, the systems are designed to promote public reporting of quality and to encourage quality improvement. Examples include:

- ♦ The Pacific Business Group on Health (PBGH) has had a standard set of measures for managed care plans for several years, but in 1997 and 1998 they initiated a performance measurement program based on patient survey data for physician group practices that has become part of their HealthScope program.
- ♦ The three major auto companies, in collaboration with other large Southeast Michigan employers, initiated a comparative

hospital profiling project in 1996. The report included (and continues to include) information on mortality, complications, length of stay, and cost for several classes of hospital admission. The concept has been extended to reports in the Cleveland, Atlanta, and Buffalo areas.

- ♦ The states of New York and Pennsylvania have supported very detailed hospital- and physician-level cardiac surgery performance reporting systems. Each year, a report is produced that focuses on mortality rates and ranks all hospitals and surgeons performing a set minimum number of procedures. Because of the narrow focus on cardiac surgery, the performance measures include a sophisticated risk factor and severity adjustment system.
- ♦ The University of Toronto was asked in 1997 to develop a standard performance measurement set for Ontario hospitals, for use by the Province of Ontario (and the general public). The measure set was built on an existing Ontario Hospital Reporting System (financial and statistical data) and the Canadian Institute for Health Information (clinical and utilization) data sets.

The widespread adoption of Continuous Quality Improvement (CQI) principles in health care organizations through the 1990s led to a much stronger internal focus on data collection and analysis as an essential tool of management. Concepts of “*balanced scorecard*,”<sup>11</sup> “*quality compass*,”<sup>12</sup> or “*instrument panel*”<sup>13</sup> all grew out of an idea that management should be more data-driven, and specifically should be based on a set of performance measures reflecting key organizational processes related to key customer expectations. No single performance measurement system or measure set has emerged from the CQI culture for widespread use; in fact, some CQI concepts would suggest that performance measurement systems should be tailored to the priorities of local “process owners” and local customers.<sup>14</sup>

Finally, a voluntary effort to develop and test *performance measures for vertically integrated health care systems* (CRISP – Consortium for

Research on Indicators of System Performance) had the active participation of as many as 24 health systems between 1992 and 1997. The project faced a number of significant challenges in defining standard data collection procedures for organizations of widely varying structures and missions. Project participants eventually concluded that the integrated health care system was perhaps the wrong unit of analysis – too large, too diverse, and too cumbersome – and the consortium evolved into more focused projects examining specific areas of performance measurement (e.g., depression care and racial/ethnic disparities in quality of care).<sup>15</sup>

### Summary

Clearly, performance measurement in health care has evolved a great deal since the days of Florence Nightingale. The main characteristics of the current environment include:

- ♦ Continued strong interest in comparative performance information by both public and private purchasers;
- ♦ A widely accepted national performance measure set for managed care plans (HEDIS);
- ♦ No similar widely accepted single measure set for hospitals or hospital systems, but an approved list of several performance measure sets (ORYX);
- ♦ Continued difficulty with focusing measurement on “bottom-line” patient outcomes like mortality, morbidity, or functional status;
- ♦ Continued development of a science base for clinical performance measures (“evidence-based medicine”);
- ♦ Continued challenge of coordinating performance measures for external/public use with performance measures for internal management and quality improvement use;
- ♦ A significant role for health plans as agents of public and private purchasers to define and implement performance measures for hospitals and physicians. Health plans are often the implementing agency, if not the ultimate source, of many physician and hospital profiling projects.

## Domains of Performance Measurement

### Defining Performance

It is not possible to design, implement, or even discuss performance measurement systems without having defined “performance.” What is it that health care systems do that should be measured? Who uses the information and for what purpose?

Several major domains of performance measurement are common among many existing measurement systems.

### Quality of Care

This domain focuses on *the clinical content of care provided for defined groups of patients*. Although it may be possible to derive some very broad quality measures which relate to the entire population of a hospital or health care system (e.g., hospital-wide nosocomial infection rate or bed sore rate), most quality of care measures reflect *accepted patterns of treatment for specific clinical conditions*. In some cases, “accepted patterns of treatment” may be derived from expert consensus or local custom; in other cases, a more formal basis in published scientific literature or nationally recognized treatment guideline is available.

Quality of care is itself a heterogeneous measurement domain. There are an infinite number of potential quality of care measures available, and some structure has to be imposed in order to design a rational system.

Donabedian’s<sup>16</sup> categorization of quality measures and concepts into *structure, process, and outcome* has formed the conceptual foundation for quality measurement for the last 35 years.

- ◆ *Structure* refers to those characteristics of a hospital, health plan, or health care system which are relatively stable, and create the capacity or opportunity for good quality. Most measures in the JCAHO accreditation system or in the NCQA/HEDIS accreditation system are structural measures. Does the organization have a patient safety committee? Does it have a system for

reviewing patient deaths? Does it have policies for handling dangerous chemicals? Does it have a credentialing committee?

- ◆ *Process* refers to what the organization does. Most utilization measures (length of stay, ER visit rate, inpatient days per 1,000) are process measures. More clinically-specific process measures include mammography rates, childhood immunization rates, or rates of use of breast-conserving surgery. Process measures are typically selected because scientific research has shown a particular practice to be associated with favorable patient outcomes.
- ◆ *Outcome* refers to a measurable change in the health status of a person or group of people. Mortality is an obvious outcome measure, but there are many other measures of morbidity, clinical state (e.g., change in blood pressure), functional status (e.g., ability to return to work) and quality of life that have been used to assess organizational performance. Performance measures reflecting organizational “outcome” (e.g., change in market share) are not outcome measures from a quality of care perspective.

Over the past decade a combination of CQI culture, research on patient safety and medical errors, and the development of treatment guidelines and evidence-based medicine has produced a different categorization of quality measures. To relate the newer system to Donabedian’s system, it is probably useful to think of all of these as process measures such that the following breakdown is an additional layer of definition within that domain.<sup>17</sup>

- ◆ *Underuse* refers to the extent to which necessary or indicated services are not provided. Examples include: failure to prescribe beta blockers to patients after acute MI; failure to provide immunizations to young children or flu shots to older adults; failure to provide regular screening mammograms to women over 50; or failure to provide adequate doses of antidepressants to newly diagnosed depressed patients.

These measures are almost always expressed as rates, and are calculated after having first identified a target population of eligible patients or health plan members.

- ◆ *Overuse* refers to the extent to which wasteful, ineffective, or unproven tests and procedures are provided to patients who don't need them. Examples include: MRIs or x-rays for patients with new, uncomplicated low back pain; antibiotics for patients with viral infections; elective surgical procedures for patients without clear indications; or use of expensive patented medications when an equivalent generic is available.
- ◆ *Misuse* refers to the inappropriate use of procedures that would be beneficial if properly applied. Most instances of "medical errors" fall in this category – overdose of medications, wrong-limb surgery, failure to follow up on abnormal test results, or adverse drug interactions.

Most patient safety initiatives taken up in response to the 1999 IOM report on medical errors<sup>18</sup> have adopted the underuse-overuse-misuse terminology, so that any new performance measurement initiatives designed to focus on patient safety issues will almost inevitably use these terms to categorize sets of measures.

Within any system for measuring quality of care, there is an additional need to prioritize measurement activities. While it may be possible to identify very precise, clearly interpretable process of care measures that reflect either over- or under-use of a treatment for a rare clinical condition, there may be little value in doing so. The effort required to gather, analyze and use the data to support QI activities may only benefit a handful of patients. The same effort applied to a more common condition might provide a much larger benefit. Selection of quality of care measures for inclusion in performance measurement systems is usually driven by:

- ◆ Prevalence of the clinical condition;
- ◆ Known incidence or severity of quality problems for that condition;

- ◆ Feasibility of collecting key data elements;
- ◆ Clarity of interpretation of performance measures (i.e., either more or less of the thing must be clearly better);
- ◆ Ability of the organization to actually use the data to improve processes or outcomes;
- ◆ Importance of the condition or quality problem to purchasers, the general public, or significant stakeholder groups.

### Utilization/Cost/Efficiency

Although the term "quality" can be used to cover almost any domain, there is usually a distinction made between performance measures that focus on evidence-based clinical care and performance measures that reflect administrative or organizational efficiency. In general, utilization measures are *measures of clinical activity that focus on activity per se, rather than whether the "right" activities are being performed.*

Length of inpatient stay is an interesting example of an efficiency measure that is not interpretable as a quality measure (even though it is included in many comparative "quality" reports). Lengths of stay have been dropping since at least the advent of the DRG payment system in 1984.<sup>19</sup> DRGs provided one clear incentive to hospitals to reduce length of stay, and managed care organizations that pay hospitals on a per-diem basis have continued the pressure for shorter stays.

We have reached a point where stays may be as short as can be medically justified; any further reductions may begin to compromise patient outcomes or at least add significantly to patient and family burden.<sup>20</sup> In this context, is a shorter LOS in one hospital a measure of better quality, or worse quality? We would argue that it is neither. It is a measure of utilization that can be interpreted in that context only, unless it is supplemented by additional information on readmission rates, infection rates, or patient satisfaction scores.

A given measure may be either a quality of care, or a utilization measure with only minor changes in specification. For example, a count of x-rays over a period of time for an enrolled

population is a utilization measure, either for the health plan or perhaps for a radiology department. A count of x-rays in a population of patients with new, uncomplicated low back pain is a quality measure, though, because there is a clear evidence-based guideline stating that these tests are not useful in this population.

There are a number of utilization measures that are already in widespread use. They include:

- ◆ Inpatient days per 1,000 or admissions per 1,000 for defined populations (e.g., managed care plan members);
- ◆ Length of stay;
- ◆ Cost or charge per admission (or adjusted admission);
- ◆ Bed occupancy rate;
- ◆ Cost per member per month;
- ◆ ER visit rate.

Hospital readmission rates are another interesting “hybrid” measure, since they may reflect either quality or utilization or both. Unexpected readmissions, particularly for the same clinical condition and to the same hospital, are almost always interpretable as a measure of poor quality, but most data systems do not distinguish between unexpected and expected readmissions (for example, for intensive chemotherapy). Readmissions after a fixed time period, for a different clinical conditions, or to a different facility may, or may not reflect quality problems, but they do represent cost to an insurer.

*Most cost measures are almost impossible to compare directly from hospital to hospital or health system to health system.* No two accounting systems are exactly the same, and few hospitals have cost accounting systems that allow precise, externally comparable calculations of costs for admissions or smaller units of service.<sup>21</sup> Purchasers and insurers may work backward from their charges to calculate a “cost” (to them) per admission, but all parties typically recognize that this figure is a very complex blend of true underlying cost, accounting conventions, negotiated discounts, and varying bundles of services. Nevertheless,

comparative cost figures are included in many regional hospital “report cards” and are used by purchasers and insurers to select organizations to include in their provider networks.

### Satisfaction

Satisfaction is an accepted part of the quality domain (usually viewed as an outcome measure), but is often handled separately (e.g., adverse event rates handled by the QA or QI department; satisfaction handled by the Marketing department). There are a variety of standardized satisfaction surveys available and in use; many provided by private vendors allow comparison with other hospital or health system clients.

For purposes of this paper, a distinction should be made between satisfaction surveys and reports of care surveys.

Satisfaction surveys are just that – patient or family reports of the extent to which they were satisfied with various aspects of a health care encounter. Scores are recorded on 1-5, or 1-10, or perhaps even 1-100 scales, with one end representing extremely high and the other end representing extremely low satisfaction. Managers may use scores for individual items or related clusters of items to monitor performance.

Reports of care are similar to satisfaction surveys in some ways, but different in other important ways. Reports of care may include questions on particular aspects of a health care encounter; however, instead of asking about satisfaction, they tend to focus on more concrete and objective issues. In some instances, the questions specifically ask about problems (e.g., did you have difficulty understanding the nurse’s instructions for how to take your medication?). In other instances the questions ask about time or process aspects of the experience (e.g., how long did you wait in the ER before you saw a physician?).

The Picker Institute has been at the forefront of the movement to use patient report of care surveys.<sup>22</sup> This is typically not an either/or issue; it is certainly possible to blend question styles in a single survey, or use two distinct surveys in the same survey packet.

## Financial

This domain of performance refers to familiar measures like net gain (or loss) or operating gain (or loss). These measures reflect the most aggregated analysis of the balance between revenues and expenses. Other performance measures like days in accounts receivable, days cash on hand, etc., reflect the efficiency of financial management processes. National and local accounting standards and the requirements of bond rating agencies typically provide a high level of standardization for these measures.

Other related measures include FTEs per occupied bed, cost per adjusted discharge, or revenue per RVU. These measures reflect operating efficiency in financial terms.

## Other

There is really no obvious limit to the range of concepts that might be included in a set of performance measures. Some additional concepts are worth noting here, though:

- ♦ Cultural and linguistic competence refers to the extent to which health care organizations offer services, or have written materials available, that are tailored to the special

needs of identifiable cultural or linguistic groups in their service population. Having Spanish-language handout materials available in a hospital with a large Hispanic population is an obvious example. Several measure sets for the concept of cultural and linguistic competence have been developed,<sup>23</sup> and major public purchasers (e.g., HCFA, HRSA) are considering making some of these measures a required part of contracting for provision of services in the Medicare program.<sup>24</sup>

- ♦ Community Benefit refers to the extent to which a hospital or health care organization offers valuable services to the community, usually in return for tax-exempt status. This is a very difficult area to quantify. Provision of uncompensated or under-compensated care is relatively straightforward, but quantifying the value of community education programs, career development programs, or time and effort support for community health initiatives is much more difficult. Measuring the impact of an organization on the overall health of a community is essentially impossible, but is often discussed as part of a community benefit measurement initiative.

## Health System Structure and Performance Measurement

When performance is measured, there always has to be an entity that is performing. The entity may be as small as a single clinician or as large as a comprehensive health care system. Large organizations face special challenges in designing performance measurement systems, since they have to accomplish at least two things simultaneously:

- ◆ Making performance measures at the “macro” level (i.e., whole system or large hospital) meaningful to the individuals whose behavior is reflected in the measures; and
- ◆ Integrating performance measures at the “micro” level (individual clinician, clinical area, or inpatient unit) with measures at the macro level, so that a consistent set of messages about priorities, mission, and level of performance are sent.

An integrated health care system may have several distinct levels of organization, and therefore several distinct levels of performance measurement. The same levels will not be present in all systems, but there are some that will be common to most integrated systems.

### Levels of Performance

The most “macro” of possible levels is the system itself. A *health care system* is inevitably composed of a number of separate (either legally, physically, or functionally) operating units, but there are some dimensions of performance that can reflect performance of the whole. A system’s Board of Directors would typically review these measures on a regular basis; the CEO and other senior managers would be responsible for achieving target levels.

Financial measures are the most natural measures to include at the health system level, since dollars are a unit of measurement that applies to all organizational types and all types of clinical activities. There are well-accepted methods for aggregating operating-unit financial measures up to the system level, and Boards expect to review a set of financial measures at the system level. It is more difficult to use

quality of care, patient satisfaction, or cost measures at the system level, since the measures may only make sense for a particular kind of operating unit within the system (e.g., inpatient unit, internal medicine department).

As an example of the challenge of trying to use other types of measures at the broad system level, think about trying to calculate a system-wide hospital readmission rate in a hospital with a mix of small and large medical-surgical hospitals, and comparing it with the same rate in another system having a mix of acute medical, surgical, psychiatric and rehabilitation hospitals.

The *hospital* is perhaps the next most macro unit of measurement. Financial performance measures are easily applied at this level, and there are some quality of care, satisfaction, and efficiency measures that make sense at this level. Hospital-wide nosocomial infection rates or readmission rates may be used (presumably, adjusted for hospital case mix), and satisfaction surveys are frequently conducted and analyzed at the hospital level.

If hospitals have specific roles within a larger health care system (e.g., acute medical/surgical vs. long-term care vs. mental health vs. children’s), then performance measures can be identified that focus on unique mission elements. It is not necessary that all performance measures at the hospital level be the same within an integrated system, although one would typically ask similar hospitals to use the same measure set.

For some health care systems, a *health plan* is a significant organizational component. The HEDIS measure set and the Consumer Assessment of Health Plans survey (CAHPS) are essentially national standards, and will inevitably be included in a plan-level performance measurement system. Beyond that, though, health plans will have financial

objectives which will relate to a set of performance measures, and there will be a number of efficiency measures (e.g., cost per member per month or medical loss ratio) that are unique to health plans.

Some measures are only interpretable at the level of the *inpatient unit or clinical service*. Measures derived from the APACHE III system, for example, apply only to the performance of intensive care units. There is a special Picker Institute survey for Emergency Departments. Many clinical quality of care of measures (e.g., nosocomial infection rates, 30-day readmission rates, functional outcome measures) can technically be done at the whole hospital level, but are much more clearly interpretable at a smaller unit of analysis. Change in the scores on a functional status survey, for example, may be excellent performance measures for an orthopedic surgery service but have little meaning for an Endocrinology service.

*Physician group* is another potential level of analysis, and this can mean anything from a large, multi-specialty group practice to a small primary care or single specialty group. At the larger level, broad measures of financial performance, patient satisfaction, or clinical quality of care make most sense. At the smaller

level, some clinical quality of care measures may make even more sense (e.g., rate of prescribing beta blockers after myocardial infarction makes more sense for a cardiology group than for a large multi-specialty group). For some measures, though (e.g., mammography rate in women over 50, or rate of advanced stage breast cancer at diagnosis), the denominator population may be too small at the physician group level to allow for calculation of meaningful rates.

*Individual clinicians*. “Physician profiling” projects take the concepts of performance measurement down to the smallest possible level of the organization. Many clinical measures cannot be calculated at this level because of small sample size issues, and there are significant questions about how to adjust clinical measures for differences across physicians in case mix. General patient satisfaction measures, and some utilization measures (again, with adequate adjustment) can be used at the individual clinician level.

In general, the smaller the unit of measurement and analysis, the more likely it is that individuals will see their own behavior reflected in the measurement reports and have some incentive to improve on those reports.

## Macro-Level Performance: Purchasers, Plans, and Populations

Macro-level measures reflect either whole-system or major system unit performance (e.g., hospitals) and serve somewhat different purposes than “micro-level” (e.g., individual clinician or small group) measures. They are typically used by either health plan leaders or external purchasers for:

- ◆ Assessing organizational performance against key strategic objectives;
- ◆ Determining executive and management incentive compensation;
- ◆ Making decisions about capital allocation;
- ◆ Setting strategic planning goals and direction;
- ◆ Interacting with regulatory and accreditation agencies or bond rating agencies;
- ◆ Comparing performance of similar operating units (e.g., hospitals or nursing homes) within large systems;
- ◆ Aligning operating unit goals and priorities with overall system goals and priorities;
- ◆ Aligning system priorities with purchaser priorities;
- ◆ Setting system-wide priorities for QI and CQI initiatives.

Macro-measures of performance are typically not used for:

- ◆ Assessing effectiveness of individual CQI initiatives;
- ◆ Calculating incentive compensation for individual clinicians or first-line managers (part of the incentive may be linked to overall system or large operating unit performance, but individual “line workers” are too far removed from overall system performance to have incentives pay calculated completely at that level.)
- ◆ Assessing impact of local technology enhancements or other types of capital improvements;
- ◆ Interacting with small local purchasers, community groups, or other stakeholder groups;

- ◆ Evaluating impact of local innovations, demonstration projects, or community-level initiatives where the system is only one of several health care organizations in the community.

### Reporting Frequency

Because “macro” performance measures are most often used by management oversight groups (e.g., Boards of Trustees) and purchasers, they are typically oriented at relatively stable aspects of system performance. They may be reported on a quarterly basis to senior managers or Boards, and on an annual basis to purchasers.

In fact, purchasers who use comparative performance measures to make decisions about future contracting relationships are ill-served by measures that change too frequently. If Hospital A is going to be chosen over Hospital B as a preferred cardiac surgery provider, it is essential that the better performance of Hospital A over Hospital B this year be repeated next year. If the key performance measures are changing so rapidly that monthly or quarterly reports are required, it is unlikely that they will be a good basis for long-term purchasing decisions. Financial performance measures are an exception, where monthly or quarterly analyses are necessary to monitor potentially significant fluctuations.

Some efficiency or quality measures may represent other exceptions to the pattern of annual reporting. An annual rate of nosocomial infections would certainly miss a sudden outbreak of infections in a single unit related to a new source of infection. An annual analysis of staffing levels per discharge would not allow for careful management of staffing to reflect seasonal variations or even weekly variations in patient volumes in different units.

Ultimately, reporting frequency for performance measures will depend on the needs of the information stakeholders to take action, and the extent to which the underlying phenomena are stable vs. variable.

## Domains and Examples

All the major performance measure domains discussed earlier can be found in existing “macro” performance measurement systems. Domains, and example performance measures include:

### Quality of Care

- ◆ Hospital-level mortality, complication, and infection rates
- ◆ Rates of specific medical errors or other patient safety issues
- ◆ Unexpected return to surgery
- ◆ HEDIS Effectiveness of Care Measures (for health plans or defined “member” populations)
  - Mammography Rates
  - Childhood Immunization Rates
  - Influenza Vaccination Rates
  - Rates of Glycosylated Hemoglobin Testing for Diabetics
  - Adequacy of Follow-up for Antidepressant Treatment
  - Beta Blocker after Acute MI
  - Prenatal Care in First Trimester
- ◆ Low Birthweight or Pre-term Birth Rate
- ◆ Five-year survival rates for specific cancers

### Utilization/Cost/Efficiency

- ◆ Bed Occupancy Rate
- ◆ Length of Stay
- ◆ Admissions per 1,000 members
- ◆ Bed Days per 1,000 members
- ◆ Cost per adjusted discharge
- ◆ FTEs per bed or per discharge

### Satisfaction / Reports of Care

- ◆ Patient-Reported Satisfaction
  - Technical Quality of Care
  - Communication/Information
  - Caring/Compassion
  - Wait Times

- Ease of Access
- Appearance of Facilities
- Parking/Food/Other Services
- Control of Pain or Other Symptoms
- Expected Results Achieved
- ◆ Reports of Care
  - Wait Times
  - Problems in Communication
  - Consistent Messages from Multiple Providers
  - Coordination of Care
  - Involvement of Family and Friends
  - Respect for Values and Preferences

### Financial Performance

- ◆ Operating Revenues
- ◆ Operating Expenses
- ◆ Total Margin
- ◆ Operating Margin
- ◆ Days Cash on Hand
- ◆ Revenue per Physician FTE
- ◆ Pharmacy Cost

### Other

- ◆ Cultural and Linguistic Competence
  - Availability of foreign-language written materials
  - Availability and ease of use of translation services
  - Number and scope of cultural competence training program
  - Provider mix reflective of community(ies) served
  - Governing board and management staff reflective of community(ies) served
- ◆ Community Benefit
  - Uncompensated Care
  - Care Provided in Public Programs (e.g., Medicaid)
  - Numbers Served in Free Clinical Service Programs (e.g., blood pressure screening, immunizations)

The above collection of domains, and measures within domains is intended to be illustrative of a macro-level system performance measure set. It might constitute a reasonable set of performance measures for many systems. However, no such stock set of measures should be adopted directly; rather, measurement sets are better when they are *carefully tailored to important system goals, mission, or priorities, or intentionally integrated with performance measures at smaller units in the system.*

Creating a performance measurement system with these characteristics requires going beyond simply plugging measures into broad conceptual categories. It involves the creation of a system that is:

- ◆ Both stable and dynamic. Some concepts and measures must be kept constant over time, so that trends can be analyzed and results of various performance enhancement initiatives assessed. On the other hand, the system must be capable of retiring measures that are no longer useful, or adding measures that reflect new priorities.
- ◆ Responsive to interests of external stakeholders. A system preference for Satisfaction Survey A is difficult to sustain in an environment where the major purchasing coalition requires all hospitals in the area to use Report of Care Survey B. It may not be wise to abandon Survey A, but it may be impossible to refuse to implement survey B. The ideal solution is to find a way to use information from both surveys in a complementary fashion.
- ◆ Conceptually and technically integrated through various levels of the system. Ideally, performance measures at smaller units of analysis in the system are “nested” within measures at larger levels so that they serve to explain variation in the more macro levels by providing more precise detail (e.g., hospital-level mediocre scores on satisfaction related to doctor-patient communication explained by department-level patient reports of care scores in communication unusually low for Cardiology and Surgery.)
- ◆ Closely linked to feasible changes in behavior. Measures that seem conceptually correct are not useful if they never change, regardless of best efforts by clinicians and administrators, or if they change dramatically in response to uncontrollable external factors. Infant mortality rates, for example, are a poor choice for performance measure, because they rarely are influenced by quality improvement initiatives in prenatal care.<sup>25</sup> Process measures are typically better than outcome measures from this perspective.

## Micro-Level Performance: Clinicians, Teams, Departments

Micro-level measures are typically used by individual clinicians, teams and departments for:

- ◆ Assessing individual or small group-level performance against best practices or key organizational objectives,
- ◆ Evaluating the impact of local process innovations and/or guidelines,
- ◆ Locally organizing and displaying information to improve knowledge access,
- ◆ Improving the management of patients presenting for care,
- ◆ Improving communication among providers, and
- ◆ Reducing variability in clinical care.

In contrast to their macro-performance counterparts, micro-performance measures tend to be *fluid, frequently reported measures oriented around specific clinical contexts*. Change in these measures is to be expected, since their *raison d'être* is to *invite, encourage and support ongoing improvement at the work-unit level*.

### Domains and Examples

Domains, and examples of micro-performance measures include:

#### Quality of Care

- ◆ Surgeon- or procedure-specific infection rates;
- ◆ Site-, section-, or provider-specific rates of prevention and screening (such as childhood immunization or colon cancer screening);
- ◆ Adequacy of lipid management for secondary prevention of coronary-artery disease (CAD);
- ◆ Identification of patients who vary from standard of care (e.g., recently discharged

CAD patients who have not received a follow-up office visit with a cardiologist).

#### Utilization/Cost/Efficiency

- ◆ Appropriate use of radiology and specialty referrals for acute low back pain,
- ◆ Cost of phone triage guideline to treat uncomplicated UTI.

#### Satisfaction and Reports of Care

- ◆ Site-, section-, and provider-specific reports of patient satisfaction.

### High-resolution

Note that many of these measures are merely macro-performance measures viewed with a higher-powered lens. For example, hospital-level infection rates viewed at high-resolution reveal surgeon- and procedure-specific infection results. The concepts may be equivalent, but differences in measurement specificity serve different purposes.

Macro-level measures are usually retrospective summaries of results. They address the question “how well did we do?” and are often viewed in the context of institutional performance as a whole. Communications around these measures tend to be oriented toward the interests of external stakeholders.

Micro-level measures, on the other hand, encourage a prospective focus on how the individual components of an organization are working together to produce a macro-result, as in “how well is my group doing now?” and “where do we need to focus our efforts in order to do better?” The context for these measures tends to be specific to smaller work-units, interested stakeholders are internal, and the motivation for change is local.

## Systems of Measurement

Performance measures, regardless of their level, should be coordinated, or integrated to form a coherent system of measurement. A well-integrated measurement system might include features such as:

- ♦ Selection of both macro- as well as micro-measures for high-priority clinical conditions (e.g., cancer care), target populations (e.g., women over age 50) or particular areas of organizational focus (e.g., process efficiency) to reflect shared values throughout the system;
- ♦ Use of identical definitions of key terms and measures for similar concepts (i.e., all medical/surgical hospitals in the system use the same inpatient satisfaction survey and present results in the same way);
- ♦ Use of comparable diagnostic and procedure codes to identify target populations and services provided;
- ♦ Use of the same financial data bases and accounting conventions to measure costs and revenues;
- ♦ High degree of sharing and compatibility of specific data elements across system units (e.g., patient identifiers, work-unit identifiers, results reporting systems, and financial databases from which performance measures for many different operating units can be derived).

This integration is easier said than done. Many hospitals have “legacy” information systems which were present before the hospital joined a multi-hospital system, or which have been built over time as stand-alone applications. Physician group practices may have medical records systems and patient identifiers which are not easily integrated with hospital medical records or electronic results reporting systems. Similarly, health plans may record individuals as members of a “subscriber unit” with the same identifier for all members of the same household, with that identifier being different from any medical record number for any of those individuals in any of the member clinical facilities.

It is, in fact, more common in practice to find fragmented, rather than integrated data systems in health care organizations. It is unusual for a large multi-specialty group practice to have integrated medical record, appointment scheduling and billing systems. It is unusual for all hospitals in a multi-hospital system to share common patient registration, order entry, or ancillary systems. It is also unusual to have hospice, home health, rehabilitation, or social work units share information systems with other health care operating units.

Note, though, that all health care organization *have* information systems. These systems may range from the state-of-the-art to the archaic. They may be more- or less-integrated; paper, electronic, or both. Measurement difficulties arise not because of a lack of data; rather, they tend to revolve around a lack of ability to coalesce those data,<sup>25</sup> to create a support system in the form of information retrieval, display and measurement tools which allow data elements to be pulled together efficiently in order to inform appropriate action. And creating such tools may be less a matter of technology than it is of determination, cooperation, leadership, common sense and creative thinking about the resources that are there already.<sup>26</sup>

### Case Vignette: Leveraging Existing Information Systems

Virginia Mason Health System is a vertically integrated hospital and multi-specialty group practice of almost 400 physicians located in downtown Seattle, with suburban clinics scattered throughout Washington’s Puget Sound region. Outpatient visits at Virginia Mason number upwards of 1 million, and the 336 bed hospital admits about 16,000 inpatients each year. Patient records at Virginia Mason are paper-based, supplemented with electronic results reporting for laboratory and radiology services.

In recent years Virginia Mason has seen consistent growth in the level of resources and interest given to quality improvement, practice guidelines and evidence-based medicine

throughout the organization. Information is the foundation for all of these efforts; however, Virginia Mason's information systems were not designed with data-based practice in mind. In fact, information systems at Virginia Mason were not designed at all; rather, they evolved over time in grass-roots fashion, much as health care information systems across the country have done. The result is a concatenation of largely stand-alone, legacy systems organized around the ability to retrieve information on individuals, one patient at a time.

Prior to 1998 Virginia Mason had little ability to reliably identify clinic populations of any chronic disease group. Their early efforts to use data systems to select guidelines topics, for example, centered around "mining" billing data, specifically, analyzing reams of charge data in a search for high volume, high cost, high variation procedures. Patient populations were identified by DRGs or primary diagnosis codes, charges were used as a gross proxy for costs, and length of stay was the most commonly used indicator of variation.

It wasn't until 1999 that Virginia Mason began developing a focused system of measurement that would make routine use of non-dollar data captured as part of everyday business—such as patient demographics,

ancillary services, specialty referrals, inpatient prescriptions and secondary diagnoses—to address clinical- and clinical quality-related data needs.

Fast forward two years, to first quarter 2001. With no new investment in technology Virginia Mason enjoys standardized, clinic-wide key indicators of performance which are updated and reported regularly at the patient, provider, site and organizational levels. Inpatient, outpatient and staff satisfaction data are combined and reported in a way that allows a "360° view" of satisfaction at the level of the institution as well as "drill-down" views of satisfaction for individual work groups. Condition-specific reports pertaining to inpatient, outpatient and chronic disease populations are commonplace. And selected data from existing computer systems have been integrated horizontally to create an electronic storehouse of information capable of supporting primary care providers in their daily clinical practice.

Like many health care organizations, Virginia Mason does not yet have a viable, electronic clinical information system. Their data stream relies on their ability to leverage existing sources of information and foster a culture of measurement-based thinking throughout the organization.

## Leveraging Existing Systems

Much of the discussion that follows draws upon Virginia Mason's experience to suggest ways in which other organizations might begin to explore similar solutions. If the reader notes a dearth of peer-reviewed literature cited in support of these suggestions, it is because few such sources could be found. In fact, a recent review article of several hundred publications from major health care and management journals concludes that:<sup>27</sup>

- ◆ relatively few studies on topics such as quality and performance management, internal systems or managerial functions such as performance have been published;
- ◆ there is a notable lack of published studies on the role and contribution of managers and governance to the effectiveness and efficiency of health care organizations; and
- ◆ most of the health care management research published is based on questions that are expenditure-related rather than value-driven.

In this review, authors Kovner and colleagues specifically call out what they believe to be a barrier to greater investment in performance measurement: "Healthcare managers," they write, "have not been trained to make strategic decisions based on evidence as have those with specialized training in research methods" (p. 6).

To address this, Kovner et al. challenge managers to combine skills with clinicians, educators, and researchers with experience in database design and analysis to create organizational systems of measurement. It is in the interest of this challenge that we offer the following ideas about how to begin leveraging the kind of information commonly available in health care organizations.

### Consider starting by addressing the needs of less demanding customers.

In the face of what can be a dizzying, often pent-up demand for complex information it is easy to miss simpler opportunities to apply existing data to clinical problems. Counting the number of office visits prompted by

uncomplicated dysuria, for example, may on the surface seem trivial compared with the need to measure, say, how well hyperlipidemia is being managed in patients at high risk of a coronary event. But of the two, counting dysuria visits may be the more do-able, and doing it well may lead to a next step in learning.

### Identify information needs.

Essential elements of any performance measurement project include the answers to key questions, such as: what is the goal of the project? What measures are needed to achieve that goal? How much information is needed, and how must it be processed to achieve the goal? Which data exist already in electronic form, who can access them, and when? Which data do we need to collect, how will we collect them, and how will they be transferred to a computer? Finally, assuming we have these data, how will we use them to influence subsequent actions? These questions may seem obvious, but they can represent a critical shift in mindset for organizations unaccustomed to explicitly data-driven decision making.

### Develop clear, operational definitions.

The world of performance indicators is filled with verbal shorthand that can impede measurement. For example, we often speak of "mammography screening rates" but less often articulate the measure behind the moniker, which might be something like "the percent of women age 52-69 seen at least once in the previous 12 months by a primary care provider or gynecologist in Health System X who received one or more mammograms within the system in the previous two years."

Indicator-specific strategies for approaching data—including diagnosis and/or procedure codes, caveats, inclusion and exclusion criteria, assumptions, etc.—can support measurement activities in a number of different ways. Well-specified strategies for approaching data lend themselves to critical review before data are pulled; this in turn encourages stakeholders to

agree on the question and method in advance of receiving the results. Such agreement helps develop data credibility, and at the same time allows room for “intelligent mistakes” which no one may have anticipated prior to analysis. Indeed, a well-articulated measure facilitates correction of those mistakes, since often only a small adjustment—adding an additional ICD-9 diagnosis code or excluding another—is necessary to address the issue.

Clear, well-documented, operational definitions openly invite stakeholders to make constructive suggestions for improvement, since those who use the data know explicitly whether the measure meets their needs, or how it falls short. Performance tracking, too, is enhanced by written definitions since they embed consistency in measurements that will be compared over time.

### Get to know your data.

Most, if not all health systems have some form of electronic, administrative (billing) system, and these have a long history of being extended to enable retrospective analysis of at least basic clinical, or performance-related questions. It can be beneficial to find one or more people in the organization who are familiar with this system, and work with them not to perform the analyses, but *to extract the data which will be used for the analyses*. This is important, since hands-on exposure to raw data can be invaluable for developing an understanding of the limits and opportunities inherent in a given data source.

### Borrow the clinical trial concept of the “ideal patient.”

Specific inclusion and exclusion criteria are often used in randomized clinical trials to limit the number of confounding factors that could obscure a study’s results. The aim is to identify the “ideal patient,” that is, the patient in whom the investigators would expect to see clearly the effect of an intervention, if indeed such an effect exists.

This concept can be applied in the realm of performance measurement as well. If your concern is reducing the extent to which

antibiotics are prescribed in connection with a diagnosis of sinusitis, for example, you may want to limit measurement to those patients receiving a diagnosis of sinusitis, and sinusitis alone at a particular office visit. Certainly these will not represent all patients diagnosed with sinusitis in your health system; rather, they will represent those patients least likely to have a non-sinusitis related reason for using an antibiotic.

The underlying assumption is that the clinic processes in place to facilitate consistent treatment of uncomplicated sinusitis will, as a matter of course, support similar consistency in the treatment of patients presenting with a mixed clinical picture. Conversely, if the system isn’t operating according to plan for “ideal” patients, then the same is likely to be true for other patients as well.

### Tap multiple data streams.

It is not unusual to find “information silos” scattered throughout a single healthcare organization, since systems of gathering information have often evolved independently over time, in response to a variety of needs and circumstances. Though administrative billing systems are often tapped for information to support performance measurement, chances are they are not the only available sources of data in a medical center.

Most ancillary services, such as lab, radiology and pharmacy, will maintain automated records, at least to some degree. Specialty databases, such as tumor registries, centralized transplant records, or surgery-staging systems may be rich sources of information. Sometimes parallel, “shadow file” systems maintained at the local level will store data electronically which would otherwise be available only through paper chart review. Finally, many medical centers participate in ongoing national registry projects, such as the National Registry for Myocardial Infarction (NRFMI), or national infection reporting initiatives such as those sponsored by the Centers for Disease Control (CDC), and these, too, may prove valuable stores of data.

Hands-on experience with the workings of different systems may suggest ways in which selected data could be extracted and cross-checked for convergence (data credibility is enhanced whenever two or more independent data sources produce equivalent results). Data from different sources may also be used in concert to enhance the specificity of performance measures, or even combined reliably into a common pool.

An example of the latter is a quarterly, provider-specific report which combines billing data containing office visit dates with actual test dates and results (e.g., microalbumin and HbA1c) obtained from electronic laboratory records for all diabetic patients seen during the prior three months.

Selected performance measures derived from these data might be:

- ◆ Percent of diabetic patients seen during the last quarter who had HbA1c measured at least once within the 365 days prior to that visit;
- ◆ Percent of diabetic patients seen during the last quarter with HbA1c levels of 9.5% or below (HEDIS standard);
- ◆ Percent of diabetic patients seen during the last quarter who had microalbumin measured at least once within the 365 days prior to that visit.

In the spirit of measurement evolution, next steps for this work could include:

- ◆ flagging those patients known to have compromised renal function (which obviates the need for microalbumin testing);
- ◆ adding additional detail, such as date/results of last lipids panel, or date of last screening eye exam;
- ◆ automating the reporting function such that feedback could be given to providers more frequently than four times per year; or
- ◆ creating a summary measure capable of tracking and reporting macro- as well as micro-level (e.g., by clinic site) performance on these dimensions over time.

### Consider the source.

Much of the data available in administrative systems are, at best, proxies for clinical activity. Data from different streams generally exist for different purposes, which themselves will influence the subsequent utility of the stored information. Thus, while it is conceptually appealing to think that ICD-9 codes, billing codes, lab work or even pharmacy data may be used to flag the presence of diabetes,<sup>28</sup> for example, not all of these data are created equal.

To illustrate, some billing systems require that all invoices be associated with a specific diagnosis or procedure code. A laboratory test ordered in an effort to rule-out diabetes, then, may be processed by the lab in association with a (differential) diagnosis of diabetes. Under these circumstances it would be a mistake to attempt to identify diabetic patients using the diagnosis codes associated with laboratory tests. Such a count would overstate the size of the diabetic population, since many of the tests could be expected to have, in fact, ruled out the condition.

Similarly, childhood immunization data are usually captured in electronic form for billing, not clinical purposes. Thus, there is a risk inherent in using billing records of immunization to infer immunization performance, since an inoculation may be given but inadvertently omitted from a billing slip, planned but later withheld and subsequently billed by mistake, or simply billed in error with a subsequent reversal of the charge.

A macro-measure of performance alone will not be sensitive enough to shed light on such distinctions. Stakeholders who doubt a reported level of performance have little recourse short of re-doing the analysis themselves. Similarly, those whose goal it is to improve performance may not know where to focus their improvement efforts. Micro-measures, on the other hand, tend to report measures in terms with which individual work-units can relate.

Such work-unit level “reality checks” can be enlightening. Upon receiving a performance report of immunization status for eligible

children in her practice, one Virginia Mason physician took exception, and set out to prove the data wrong. By chart review she demonstrated that these children had, in fact, been immunized. She also discovered that the immunizations they had been given had never been submitted for billing. Closer examination revealed a process issue at this particular clinic which caused many immunizations to be omitted from billing. Thus, the administrative data were correct insofar as they represented immunizations billed. The provider was also correct, insofar as she demonstrated clinical delivery of immunizations. It took the reality check that came between them to get to the root of the measurement problem (and identify the system issue that was—invisibly!—impacting the organization’s revenue stream).

### Beware the upper limits of a data source.

A Virginia Mason team once attempted to use billing data to identify a cohort of patients newly-diagnosed with hypertension. The intent was to conduct subsequent medical record review to determine which first-line therapies were being prescribed for the condition. Since an electronic search for newly diagnosed patients could not be done directly (ICD-9 diagnosis codes do not distinguish between newly diagnosed, and long-standing hypertension), the team crafted detailed specifications by which to identify through billing records those patients most likely to be newly diagnosed—...patients over age 40 who received a diagnosis of hypertension at an office visit occurring during the last six months, but who had not received the diagnosis in the three years prior to the visit.... Upon chart review, however, the team discovered that the strategy had failed.

The billing system, the team later realized, records as many as three diagnoses for any particular office visit, but no more. Many of the patients identified through billing data as “likely” first-time hypertensives were, in fact, suffering from *chronic* hypertension as well as any number of other comorbid conditions, such as congestive heart failure, cancer and renal disease. That hypertension appeared on their

(limited) list of billing diagnoses after a hiatus of many years was testimony to their therapeutic progress with more clinically pressing concerns, not indicative of the previous *absence* of hypertension.

### Supplement data known to be inadequate.

Sometimes data are known to be inadequate; for example, internal billing data which cannot capture things like cervical cancer screening (Pap) tests performed outside of the organization. One solution to this problem is to conduct a planned, annual chart review of randomly selected, eligible female patients lacking billing evidence of a current Pap test to determine what proportion of them in fact are current once Pap tests given outside of the health system are considered. This estimate may then be used as an explicit adjuster to the performance measure calculated from administrative data alone.

### Make judicious use of resource-intensive methods.

Prospective data collection and retrospective chart review remain the gold standards for acquiring data in health systems lacking an electronic clinical information system, but both methods tend to be resource intensive. Judicious use of these tools, however, can be invaluable.

An example of this is the indicator “percent reduction in potential medication errors,” which must be supported by data collected prospectively. Such measurement can be made efficient by tailoring it to the needs of the work. One way of doing this is to develop tools to maximize data consistency and timeliness, and minimize the burden of data collection and analysis. An example of such a tool—a simply designed Microsoft ACCESS™ screen that helps catalog and understand the sources of potential error in our medication orders—is shown in Figure 1. Data are collected quickly, and in a form immediately available to the work group for analysis, since reports pre-programmed into the application are available on demand as soon as data have been entered into the system.

Retrospective chart review can be used as a primary source of data not otherwise available, but also as a backup method for addressing issues of “data disbelief,” such as colon cancer screening rates drawn from billing data which are believed by stakeholders to be too low. Rather than simply abandoning administrative data sources as unreliable for this indicator, focused chart review may be used to calculate the sensitivity and specificity of the administrative data used to generate the measure. This gives stakeholders a basis upon which to decide whether subsequent measurements made using billing data are likely to be accurate enough to meet their needs.

### Establish goals and accountabilities.

We have already discussed the need to establish operational definitions and clearly specified rules of measurement. Remaining is to identify goals and accountabilities for performance that will motivate progress.

External benchmarks for performance can be set using national results, such as those from NCQA, wherever possible, though the proliferation of evidence-based “best practice” information tends to be limited in health care.<sup>27</sup> Barring this, performance can be benchmarked against regionally available data or previous internal results.

Some health systems might approach the setting of minimum, target and stretch goals for performance quantitatively, by calculating statistical confidence limits around the benchmark identified as the “minimum standard.” From there, targets could be set at the level necessary to produce a statistically significant ( $p < .05$ ) improvement over the minimum standard, with stretch goals set at a  $p < .01$  level of improvement.

Other systems might be more comfortable taking a qualitative approach, striving for what they would regard as clinically, or operationally meaningful improvements in performance, regardless of statistical significance. The important point, though, is that goals and accountabilities around performance are declared and adopted.

### Improve Incrementally

Some may claim that all clinical measurement problems will be resolved eventually by increasingly sophisticated information technology, but in truth it is unrealistic to expect any measurement system to be “plug and play,” capable of providing turnkey answers to complicated clinical and management questions. Many of these problems require the painstaking development of mixed solutions, especially since the environments in which clinical tasks are carried out are constantly changing.<sup>29,30</sup> In this regard, a focus on the very people-based task of building incremental prototypes of solutions, rather than ultimate answers, may prevent both measurement and technology from becoming obsolete, or worse, ends in themselves.

By their nature incremental solutions have the distinct advantage of flexibility, encouraging stakeholders to assess what’s been learned at every step and then define for themselves the next level of need. In this way, clinical and administrative stakeholders are ensuring that the systems and measures developed remain relevant. Indeed, by adopting key indicators, building common definitions, analyzing information and using data displays—even paper pro forma displays—stakeholders are improving the likelihood that technological innovations, when they do arrive, will quickly be used to potential.

## Specific System Examples

In this section we provide a number of specific examples of health care systems which have done a particularly effective or innovative job of creating performance measurement systems that link different levels of analysis within the system.

### Veterans Health Administration

The VA is one of the nation's largest and oldest vertically integrated health care systems. It was originally designed around an acute hospital and rehab facility core, but has grown to include primary care centers, nursing homes, specialized mental health facilities, and other components of a truly integrated system.

In the past 20 years, the VA has made important advances in the development of clinical and administrative information systems; these systems have allowed the VA to move to a leadership position in the area of performance measurement.

One of the key organizational changes for the VA has been the reorganization of facilities and staff into Veterans Integrated Service Networks (VISNs). This system has been in place since 1995, and replaced the older system of districts and regions that had been developed mainly around issues of hospital service areas.

There are 22 VISNs; each one includes 7-10 hospitals, 25-30 ambulatory care facilities, 4-7 nursing homes, one or two domiciliaries, and a number of counseling centers (ref). Each VISN has a veteran population in its services area of approximately 200,000. The VISN serves as the primary unit for budget management, program coordination, and quality improvement. The VISN model of organization has been accompanied by a stronger emphasis on primary care and continuity of care, so it is reasonable to think of the VISNs as having an "enrolled population" for whom the VISN provides a continuum of needed health care services.

A Performance Management Program was developed in 1995 to serve as a framework for a variety of quality improvement and management accountability initiatives. The system has a number of measures of both processes and outcomes of care organized around 10 major dimensions. The measures reflect care provided at multiple sites and levels, not just inpatient care. The PMP has supported a number of more specific initiatives and programs:

- ◆ The VA's National Surgical Quality Improvement Program has been a model for large-scale surgical quality improvement efforts. It uses risk- and severity-adjusted functional status and mortality data to compare hospitals, track trends over time, and promote improvement toward benchmark levels.
- ◆ Specific performance measures are used in performance contracts for managers. Managers are held accountable for achieving target levels of performance on one or more measures within specific time frames. These initiatives have been associated with major improvement in a variety of morbidity and mortality measures.
- ◆ In the ambulatory setting, development of a Prevention Index (PI) and a Chronic Disease Care Index (CDCI) has drawn clinical and management attention to a set of evidence-based performance measures that reflect key aspects of longitudinal care for veterans. The PI focuses on health education, prevention, and screening, while the CDCI focuses on care for veterans with diagnosed chronic conditions like diabetes or hypertension.
- ◆ Similarly, a Palliative Care Index has been developed that includes a set of measures on process of care and patient and family satisfaction. In all of the indices the global index measure can be used by senior administrators to track broad trends and compare facilities; the individual measure components can be used by local clinicians and managers for focused QI activities.

- ♦ The Quality Enhancement Research Initiative (QUERI) is a relatively new initiative (1998) that attempts to link clinical and health services research to quality improvement. QUERI attempts to take the latest scientific evidence on quality/best practices, incorporate it into the performance measurement system, and then use the existing accountability and management oversight mechanisms to drive performance in the desired direction.

The hierarchical administrative structure in the VA allows for a clear sending of signals throughout the system about clinical priorities, target values or benchmarks, definitions of measures, and other aspects of a system of measurement and accountability. Except possibly for HEDIS, there may be no other system of performance measurement that has a comparable level of reach and impact.

## Henry Ford Health System

The Henry Ford Health System is a not-for-profit, vertically integrated health system serving Southeast Michigan. The system includes five hospitals, the largest managed care plan in Michigan, a 1,000-member multi-specialty group practice, and a network of ambulatory care clinics, vision centers, rehabilitation and physical therapy centers, and long-term care facilities.

Performance measures at Henry Ford follow a “layered” approach, with a basic set of measures collected, reported, and applied at the system level, then related sets of performance measures collected, reported, and applied at the level of operating units (e.g., individual hospitals or health plan), smaller units within legal operating units (e.g., inpatient units of a single hospital, departments in the medical group, primary care sites), and ultimately to the level of individual clinicians.

At each of these levels, except for the individual clinician, the organizing scheme for the performance measure set is the list of current strategic priorities identified by the System

Board. This is a short list of broad concepts. In the late 1990s, for example, the four priority concepts were growth, low-cost provider, satisfaction, and clinical quality. These priorities may change somewhat over time, but efforts are made to keep the broad priority areas as stable as possible over periods spanning several years.

Within each broad area, 4-6 specific measures are selected for analysis and reporting at each level of the organization. For example, the health plan might select total cost per member per month as a measure under the concept of “low-cost provider”. At the hospital level, though, the pmpm measure doesn’t make sense, but a measure like “FTEs per adjusted admission” might be used. Measures at the system level are typically a blend of health plan, hospital, and medical group measures, since the system per se does not have a natural set of performance measures in domains other than financial.

The content of the performance measure set is driven by the “balanced scorecard” concept, where there is a conscious effort to draw measures from a set of clinical, administrative, and patient/member satisfaction domains.

The display format for the measures is a “spider diagram,” in which the four major priority areas each occupy a quadrant, the specific measures represent “spokes” radiating out from the center, and the actual level of each measure is expressed by the distance from the center along each spoke that a mark is placed. A (normalized) target value of 1.0 is assigned for each measure, and connecting those dots creates a circle. Plotting actual performance may result in dots that are inside or outside the circle, with outside being better, and inside worse than target. Overall performance, then, can be captured in a visual way that is intuitively appealing. A bigger, more regular circle of connected “actual” dots is good; a smaller, less regular circle is not as good.

Data to generate these reports at Henry Ford come from a variety of clinical and administrative data sets. There is a “Corporate Data Store” that drives many of the measures –

it is a summary of clinical encounter and patient registry data that is used for a variety of clinical and administrative analysis purposes.

Some specific points of emphasis about the Henry Ford performance measurement system:

- ♦ There is an explicit linkage between high-level system priorities, system-level measures, and performance measures throughout the rest of the organization. The unit-level spider diagrams are locally tailored variants of the system spider diagram.
- ♦ The spider diagrams at all levels include a blend of clinical, financial, and patient-reported measures to create a “balanced scorecard” approach. Cost containment efforts that succeeded but had a negative effect on patient satisfaction would show up as an obvious irregularity in the spider diagram and would not be fully satisfactory.
- ♦ The measures are used primarily for management analysis, administrator and clinician incentive systems, and Board and community oversight. The measures are not generally clinically precise (i.e., focused on a single clinical condition or procedure), so they do not naturally feed into CQI team efforts. A different set of performance measures, driven by more clinically precise data systems, typically supports CQI initiatives and other performance improvement programs. The performance measures do, though, serve to set priorities for broad QI initiatives.

Individual operating units may have performance measurement and reporting requirements that go beyond the four main areas and the spider diagrams. Health Alliance Plan (HAP), for example, is required to produce HEDIS reports and administer the CAHPS survey each year. System hospitals must participate in the ORYX system as part of their JCAHO accreditation requirements. There is some synergy across these efforts (e.g., HAP can use one or more HEDIS or CAHPS measures in its own spider diagram), but the performance measurement activities in the system are actually much larger, more complex, and less closely

connected that the related sets of spider diagrams would suggest.

## Lovelace Health System

Lovelace Health System is a vertically integrated system serving the Albuquerque area. Its historical roots are in the Lovelace Clinic, a large multi-specialty group practice. Over the past 15-20 years, the Lovelace Health Plan, the Lovelace Clinic Foundation, and an affiliation with Cigna have expanded the size, scope, and mission of the organization. Lovelace is now recognized as one of the top 10 integrated health care systems in the country.

Lovelace was one of the national leaders in the early 1990s in developing capabilities for measuring health outcomes as quality of care indicators. In the mid-1990s, Lovelace identified a set of high-priority clinical conditions (e.g., diabetes, asthma, CHF, back pain) and developed an Episodes of Care system to integrate performance measurement, quality improvement, and disease management for those conditions.

The Episodes of Care initiative has several goals. These include:

- ♦ Measurement and improvement of patient health status
- ♦ Enhanced secondary prevention to reduce complications and morbidity for patients with chronic disease
- ♦ Enhanced quality of care and reduced variation in treatment
- ♦ Better network management for non-staff model physicians.

In a typical Episodes of Care initiative, a multidisciplinary team reviews scientific data on best practices, available guidelines, and current data on Lovelace practices to identify opportunities for improvement and develop locally-tailored guidelines. A measurement system is developed for key quality indicators (e.g., rate of dilated eye exams or glycosylated hemoglobin levels for diabetes), and reports are

produced for both system leadership and individual clinicians.

Some of the measures are taken directly from standard employer-oriented performance reports, such as HEDIS. Improvements through CQI processes show up directly in better HEDIS scores.

A number of published articles have documented the positive impact of the Episodes of Care system on key quality of care measures for the target clinical conditions.<sup>31</sup>

## **Scripps Health: Critical Indicators of Enterprise Performance**

Scripps Health is an integrated health care system serving the San Diego area. The performance measurement system at Scripps is based on two organizing principles: the “dashboard” and the “balanced scorecard.” The dashboard concept suggests a set of measures useful for those who are “driving” or “steering” the system. A set of measures are available that display, in an easy to understand and concise fashion, what the system’s current status is on those measures.

Measures are chosen in a way to facilitate management decision-making and action. The balanced scorecard concept suggests that measures must be drawn from a set of clinical, administrative, patient satisfaction, and community service domains rather than all from just one or two domains.

An initial set of measures and concepts was identified in 1993, and the main features of the graphic display format were developed as part of the process of affiliating with Mercy Hospital. The selection of measures to reflect a balanced scorecard concept came in 1997 and 1998. The balanced scorecard was developed to reflect seven major domains related to Scripps’ key elements mission:

- ◆ Community health and advocacy
- ◆ Customer service and community served
- ◆ Financial

- ◆ Medical excellence, operational
- ◆ Medical excellence, clinical outcomes
- ◆ Information management
- ◆ Values, staff investment, and development

Each measure in the overall set is part of one of these seven domains. When managers seek performance information on its “intranet”, they can find individual measures organized by the seven major domains. Each domain has a number of specific measures. For example, Customer Service includes 12 specific measures drawn primarily from patient satisfaction surveys. There are 5 measures of inpatient care, two measures of outpatient surgery, two measures of outpatient treatment and testing, two measures of emergency department care, and one measure of care to the entire community.

There are a number of unique features to the Scripps Critical Indicators system that give it special value:

- ◆ The bringing together of inpatient, outpatient, ER, and other site of care measures in the same overall measure set creates a stronger sense of “systemness” than other models where the measurement sets are organized separately by operating units, regions, or levels in the system.
- ◆ The set includes a clinical outcomes component. Although health care organizations have been trying (in one way or another, for over 100 years) to use clinical outcomes as key “bottom line” performance measures, there are few successful examples of that actually being done. It is commendable that Scripps has been able to develop a set of measures and include it as one of the seven major performance domains.
- ◆ The system is organized to serve the information needs of senior management and the Board of Trustees. A different system might have evolved if it was designed for a different audience or different purpose.
- ◆ The measures are not just broad, aggregate measures with no link to daily operations or CQI activities, though. The measures were

selected to reflect sets of discrete clinical or administrative processes with their own “sub-measures”. Projects focused on small, local processes would see change most immediately in the “sub-measures”, but would also see effects in one or more dashboard measures if the effect was big and broad-reaching enough.

- ◆ Control charts are used as a preferred graphic style for presenting performance measures data. This format allows the user to see time trends, variation around average or target values, and particularly significant variations that are highlighted in color.
- ◆ The reports are available to all system employees who have access to the system’s computer system. This creates an open environment in which key performance dimensions of the system are widely known and available for use in more local strategic planning and quality improvement programs.

### **Sharp HealthCare: Sharp HealthCare’s Outcome Tracking Initiative**

Sharp HealthCare is a not-for-profit integrated health care system based in San Diego. The system includes six hospitals, three medical groups, a health plan, a home health agency, and a number of clinics and skilled nursing facilities. Sharp has approximately 10,000 employees and 5,000 affiliated physicians. There are 350,000 covered lives in capitated plans, and Sharp has approximately 25% of the health care market share in the San Diego area.

Sharp has developed a measurement system that includes more than 3,000 indicators of clinical and financial performance. These performance measures are focused on 30 disease states or clinical conditions (e.g., acute myocardial infarction, congestive heart failure, pneumonia, childbirth, diabetes). Many of the measures are analyzed and reported at the facility (i.e., hospital) level, but reports are also prepared for individual physicians within

hospital, and for individual affiliated physicians on important areas of non-hospital care (e.g., ongoing management of patients with diabetes).

There are several interesting features of the Sharp performance measurement system that are worth emulating by other organizations:

- ◆ There is a clear scientific and management basis for the structure of the measurement system. It is built around 30 diseases or clinical conditions, and most of the effort is focused on the ten conditions that involve the highest patient volumes, costs, potential quality problems, and opportunities for improvement. The reports, then, focus on patient groups and processes that both clinicians and managers find meaningful.
- ◆ A great deal of effort is expended to assure that the reports are accurate, timely, and adjusted for relevant dimensions of patient risk or severity. The reports are prepared by groups working under physician leadership, and every effort is made to assure that reports are not dismissed because of being invalid or out of date.
- ◆ Related measures for a particular condition are presented together in the same graphic format, with variances from norms or target levels highlighted. Clinical and financial measures are displayed together in many of the reports. This allows both clinicians and managers to see the same “picture” of the care of a particular condition, and to see patterns in the variances across different performance measures. (For example, if complication rates, length of stay, and costs are all better or worse than average in a given hospital, this is apparent in the report in a way that would not necessarily be so if the performance measures were on different pages or in different tables, particularly if they were divided into clinical vs. administrative and presented to different audiences.)
- ◆ When possible, run charts are presented showing change in rates over time, at monthly or quarterly intervals. This allows users of the reports to distinguish at a glance between essentially random variation and meaningful trends. It is also possible to see

the effects of CQI initiatives or other management or clinical initiatives.

- ◆ The reports are designed to “drill down” to the smallest possible unit of meaningful analysis, including the level of the individual physician. There is little opportunity for teams or individuals to dismiss reports as being relevant only to someone else.
- ◆ There is an intentional focus on measures that are “actionable”. Performance measures are either direct or indirect reflections of specific processes of care, all of which are at least theoretically modifiable. The performance measure set is also embedded in a disease management and clinical quality improvement structure that routinely connects data to action. There is a clear integration in staffing between data collection and reporting on the one hand and action on the other.
- ◆ There is stability in the clinical conditions selected, measures used, and reporting formats displayed. The purpose of this stability is to maximize users’ comfort with, and ability to use, the information. The organizers of the effort attempt to minimize surprise and bewilderment among users so that the focus of attention can be clearly on levels of performance, trends, variations among units and providers, and opportunities for improvement.

Primary contacts: Mark Slater, Ph.D., Lori Anderson, MPA.

### **Virginia Mason Health System: 360° Satisfaction**

At Virginia Mason Health System in Seattle, inpatient, outpatient and staff satisfaction surveys are conducted independently of one another. Data from each survey are integrated and reported at the level of the institution as well as in “drill-down” views of satisfaction for individual work groups. In addition, Virginia Mason reports an integrated, “360° view” of inpatient, outpatient and staff perceptions on 16 key components of satisfaction.

The system grew from formal and informal discussions with internal stakeholders, each of whom had a need to use data from the satisfaction surveys for slightly different purposes. At the same time, it was deemed important thematically that the reports be designed to embody the concept of data “rolling up” and “drilling down” to reflect performance at all levels in the organization. What was needed, then, was a means of displaying satisfaction data which would address as many of these requirements as possible.

Figure 2 shows a sample page from a Vice President-level report of staff satisfaction survey data at Virginia Mason. Staff satisfaction data are reported similarly at the director, manager, clinic, department, and individual work unit levels. (Results from other satisfaction surveys at Virginia Mason, including inpatient and outpatient satisfaction, are presented in a like fashion.)

Key features of this layout include:

- ◆ Survey items are organized into sections (e.g., “Relationship with Supervisor”) such that like-items appear together. Item groups were determined, and later confirmed using factor analysis of the survey data. These sections allow stakeholders to understand performance in the context of individual, satisfaction-related items, as well as in a broader context of conceptual dimensions that influence satisfaction.
- ◆ Performance for the organization as a whole is shown on all reports (“VMMC Overall 2000”). This gives stakeholders an internal benchmark with which to compare their own results.
- ◆ Results at each organizational level (in this case the Vice President level) are displayed over time (previous year, current year), with the difference from one year to the next called out as a positive or negative change score.
- ◆ Change scores are evaluated (by Z-test), and statistically significant differences in result from one year to the next are called out. (Pains are taken, however, to remind stakeholders that statistical significance does

not necessarily indicate meaningful clinical, or operational change.)

- ♦ The right-most column labeled “High response variation” is used to indicate items for which there were a large number of positive *and* negative responses given; that is, where results may be suffering from unevenness or inconsistency in the organization.

Beyond this, Virginia Mason combines results from independent inpatient, outpatient and staff satisfaction surveys to allow a “360° view” of the organization’s performance on 16 key components of satisfaction (see Figure 3). Results from like-items across surveys are plotted on the same performance “wheel,” inviting questions such as “are our outpatients and inpatients equally satisfied with our services?” and “do our staff self-assessments of performance related to patient care match what our patients are telling us?” The display encourages the integrated use of data sources which might otherwise be viewed and interpreted largely in isolation.

Primary contacts: Nancy Neil, Cathie Furman

### **Allina HealthCare: Collaboration in a multi-clinic system**

Allina Health System in the Twin Cities metro market is large, multi-faceted and multi-facility, geographically expansive and heterogeneous. Though Allina is integrated in the business sense, its 47 medical clinics remain largely autonomous.

Understanding that congestive heart failure (CHF) is one of the areas with the greatest potential for improved outcomes and efficiency in care management, representatives of four Allina sites—Litchfield, Hastings, Coon Rapids, and Nicollet Mall—came together a little over a year ago to explore how they might improve care for their patients with CHF.

Three to five individuals came from each participating site for these organizational meetings, which focused initially on the setting of common aims and goals. The group was led

by a medical director (internist) with collaboration from support people in staff positions for care improvement, and also included a cardiologist, who served as a content expert for the work. After these meetings, representatives from the participating sites went back to their clinics and participated in a thorough process in which physicians and other clinicians sat down together to reach consensus on standards of care and protocols for treating CHF.

One element that posed both challenges and opportunities for Allina was the diversity of the individual clinics involved. Litchfield is a rural site with 13 providers, in the western part of Minnesota, outside the Twin Cities. Hastings is a suburban site with 25 providers; Coon Rapids is a large, 80-provider medical clinic; and the Medical Arts Clinic is located near Nicollet Mall in the heart of downtown Minneapolis. Nevertheless, the guiding principle for the work has been internal collaboration, and the process itself broadly universal among the four sites. Each CHF management team was composed of generally the same constellation of professionals: a physician champion (family practitioner or internist), a care improvement coordinator or facilitator (typically a nurse), and an operations person, typically a clinic manager.

“The common aims and measures were to improve the care of patients with congestive heart failure, as measured by the percent of patients who’d had appropriate diagnostic work-up, including echocardiography, and the percent of patients who were appropriately classified according to the New York Heart Association classification scheme for heart failure,” explains Ted Loftness, M.D., medical director for Allina Medical Clinics. Those goals, Loftness says, were universal among the participating sites, though he adds that “some sites had unique goals, such as measuring the percentage of patients with living wills, or percentage of patients enrolled in a cardiac rehabilitation program.” This resulted in some natural synergies. “Sites learned from one another,” Loftness emphasizes; “there was a lot of teaching around what worked and what didn’t

work, as well as strategizing around what kinds of things engaged the providers.”

Allina encountered several interesting issues in the course of their work in CHF:

- ◆ Allina initially used claims data to try to identify patients with CHF so that they could supplement these data with chart audit and review. A key finding from this exercise was that only about one-third of their patients who had claims around CHF were appropriately classified (clinically) for the diagnosis. This raised questions around the appropriateness of coding. For example, they found that many patients presenting at Allina clinics with shortness of breath were incorrectly being given diagnosis codes related to CHF. Further investigation revealed that the problem was consistent across the organization’s clinics. The attempt at measurement, then, helped uncover the problem which, in turn, led to coding improvements.
- ◆ Pursuing the CHF performance measurement initiative helped Allina physicians reach consensus around standards of care. At the start of the project most Allina providers agreed that ACE inhibitors were effective for CHF patients, but some questioned the efficacy of beta blockers for [NYHA] Class III and IV patients. “Now there is consensus on that here,” notes Loftness.
- ◆ Part of the work of the initiative at Allina was to build a CHF patient registry at each participating clinic site, detailing patients’ diagnostic test results, selected medication regimens, NYHA classification, and other notations, such as the presence of a living will. Participants observed that the very fact of recording and measuring naturally had a strong impact on performance, and the ongoing dialogue among team members facilitated interactions between cardiologists and primary care practitioners, as well.
- ◆ With its diverse mix of clinics, the key hurdles for Allina included finding ways to translate, or transport learning between organizations within the system, and to bring people together to allow them to share.

In this case the umbrella clinic organization is meeting the challenge. They started small—with 4 out of 47 clinics participating—but managed to demonstrate improvement on self-selected parameters. Now four more clinics have joined the CHF initiative, and Allina plans to expand its disease management model to asthma, diabetes and migraine care in the near future, using the same collaborative format.

Primary contact: Maureen Schreiner

*Adapted from an April 2001 article prepared for Data Strategies and Benchmarks by Mark Hagland, entitled “Allina’s experience with CHF management—a multi-clinic/integrated health system perspective on disease management development.” Content used with permission.*

## Dean Health System

Dean Health System is an integrated health system located in Madison, Wisconsin. Its historical roots are in the Dean Clinic, a large multi-specialty group practice that has been serving patients in the Madison area since 1904. The Dean Health Plan was developed in 1983, and the Dean Foundation established in 1986. The Dean Clinic’s medical care activities evolved into the Dean Medical Center, which now includes 315 providers at 22 locations in the Madison area. The Dean Health System has a close affiliation with St. Mary’s Hospital Medical Center; a joint venture between Dean and St. Mary’s provides a management structure for 75 primary care providers in a 10-county area. This network provides a strong, stable referral base for the specialty care provided by Dean and St. Mary’s.

In 1990, Dean joined the American Group Practice Association (now AMGA) in a national “Outcomes Management” initiative. Dean committed to formal measurement of self-reported patient outcomes in several clinical areas, and created a structure that allowed the information to not only be used for aggregate performance reporting, but also for individual-

level patient care by including information in patient charts.

Outcomes Management projects were implemented in a number of clinical areas, but Dean took a particularly strong leadership role in measuring outcomes for patients with either asthma or depression.

More recently (2001), Dean has made a commitment to quality improvement projects in all 27 medical departments. Within each area, staff are expected to develop specific performance measures in three major domains:

- ◆ Medical management (i.e., quality of care);
- ◆ Practice development (business/operations); and
- ◆ Patient satisfaction.
- ◆ Examples of specific performance measures include:
  - ◆ Use of antibiotics in upper respiratory infections (goal is to reduce rates);
  - ◆ LDL levels in patients with coronary artery disease (higher rates of achieving target levels);
  - ◆ Percent of CHF patients on ACE inhibitors;
  - ◆ Standardization of cataract surgery supplies and equipment;
  - ◆ Level of pain management post-arthroscopy;
  - ◆ Percent of hip fracture patients having bone density tests (goal is to increase rates).

A number of other measures are currently under consideration in other clinical areas.

## Legacy Health System

Legacy Health System is an integrated health care system serving the Portland, OR area. There are two tertiary care hospitals, two community hospitals, a children's hospital, home health agency, research center, and a network of primary care centers. A complete range of primary care, specialty care, and special population services (e.g., seniors, women, children) are available. Legacy and its facilities have won a number of quality and community

services awards, and Legacy is the sixth-largest private sector employer in Oregon.

Legacy has a detailed performance measurement system that commits the organization as a whole, and individual units within the organization, to achieve benchmark levels of performance in a wide range of clinical and operational indicators.

The performance measurement system is built around four "dashboard" sets of indicators. The Strategic Dashboard is used by system senior management to monitor key measures at various parts of the system. Specific indicators include:

- ◆ Mortality rate
- ◆ Quality improvement
- ◆ Readmission rate
- ◆ Central venous catheter infection rates
- ◆ ORYX measures for acute care
- ◆ CABG mortality rates
- ◆ Aspirin administration within 24 hours post-MI
- ◆ Positive blood cultures in NICU
- ◆ Birth trauma among high risk women
- ◆ C-section rates
- ◆ C-section infection rates
- ◆ Surgical site nosocomial infection rates
- ◆ Stage of cancer at diagnosis
- ◆ Market share
- ◆ Cost per case

The Employee Satisfaction Dashboard includes:

- ◆ Overall job satisfaction rate
- ◆ Level of pride in working with LHS

The Customer Service Dashboard includes a set of patient satisfaction measures that can be analyzed at an aggregate system level as well as for individual hospitals. These include:

- ◆ Percent who recommend LHS to others
- ◆ Percent rating "5" (Excellent) on survey items
- ◆ Inpatient satisfaction with

- Parking
- Courtesy of staff
- Cleanliness
- Caring of staff
- Quality of meals
- Follow up education and instruction
- Pain management
- Overall satisfaction

based in QI projects can find a place in one of the dashboards as long as the area remains a high priority.

The Key Processes Dashboard assess more micro-level clinical measures (mainly process, but some outcome) that relate to performance in individual units or divisions. Examples include:

- ◆ Percent of transfusions having reactions
- ◆ Accurate performance of transfusion protocols
- ◆ Adverse drug reactions
- ◆ Medicine error severity ratio
- ◆ Autopsy rate
- ◆ Correlation between pathology reports and autopsy findings
- ◆ Employee exposures to blood and bodily fluids
- ◆ Hours of emergency room diversion
- ◆ Code response time
- ◆ Restraint use rates
- ◆ Specimen rejection rates
- ◆ Abnormal mammogram turnaround time

Other clinical measures in the visiting nurse and home health program that have the same basic structure.

This measurement system is embedded in a management system and culture that emphasizes “managing with data” as well as a focus on patient satisfaction and continuous quality improvement. There is a formal Quality Planning and Monitoring process that is followed by quality improvement teams and work groups. The quality improvement initiatives are selected through a formal priority setting process that includes the relationship between a particular problem or improvement opportunity and LHS strategic goals. Measures in the various dashboards are used to identify opportunities for improvement; some measures

## Summary Thoughts

We have presented many examples of health care systems that are engaged in innovative, successful work in the area of performance measurement. We know that many other such examples exist. These systems are actively demonstrating not only that performance measurement can be done, but that it can be done well. As we review these systems' accomplishments we notice several common threads, from which we draw the following summary thoughts.

- ◆ Interest in comparative performance information for health care organizations remains strong, both by stakeholders within, as well as outside of health care.
- ◆ There seems to be little debate that performance measures in health care should include some mix of clinical quality, patient (or member) satisfaction, efficiency, utilization and financial performance. The key questions that remain have to do with how a balanced set of measures should be selected, how they should be implemented, and how they may best be used for both external accountability and internal quality improvement purposes.
- ◆ Even as the general concept of a “balanced scorecard” gains favor, it is clear that the perfect measurement set has not yet been invented. Every measure has one or more flaws, as do data available to support them. Managers may take courage, however, in the knowledge that performance measures need not be perfect in order to be useful. As was evident throughout the case examples, a “good enough” performance measure (i.e., showing evidence of a problem when there really is one, or reflecting trend in the actual direction of change) may be all that is needed to take the organization to the next level of improvement.
- ◆ From this stems the corollary that measures tend to improve when people use the data to inform decisions that matter. The experiences of Allina HealthCare and Dean Health System, among others, illustrate that data which are inadequate to start can be cultivated when the people who use them understand the importance making sure the data are complete, accurate, and tailored to the analysis at hand.
- ◆ Even the most accurate data, though, are not useful if there is either too much, or too little “organizational distance” between the unit of analysis for the data and the unit of control for making change. Scripps Health, Virginia Mason Health System, the VA, and Henry Ford Health System, among others, have embraced this concept, each instituting a top-down, “layered” approach to performance measurement in their systems.
- ◆ Finally, performance measurement is best supported when it rests on a clear scientific and statistical foundation. Sharp Health Care, Legacy and Lovelace Health Systems all have well developed, sophisticated measurement foundations; other systems may judge their own foundations to be yet on that path. Still, while there is no doubt that professional-level research methods and statistics are sometimes useful, it is also true that basic principles and clear thinking around data issues are often all that is necessary to see one through to meaningful conclusions.

Figure 1. Example of an electronic primary data collection tool around sources of potential medication ordering errors.

**Medication Review Form**

Order ID  Rx #  Reviewer Initials  Review Date  Order Date

Order Type  Floor  Patient Initials  Rx Type  Pre-Printed Order  P-Notes

**New Entry**

**Cancel Entry**

**Return to Main Menu**

**LEGIBILITY**

- Illegible order requiring clarification or educated guess : DRUG
- Illegible order requiring clarification or educated guess : DOSE
- Illegible order requiring clarification or educated guess : SIG
- Illegible order requiring clarification or educated guess : FORM
- Illegible addressograph

**ALLERGY**

- No allergy indicated
- Medication ordered that patient has an allergy to
- Standing order with no section for indicating an allergy

**NON-STANDARD INFORMATION**

- DOSE outside PI range or VM guidelines
- SIG outside of PI range or VM guidelines
- Non-formulary drug
- Unknown Drug
- Non-standard drug abbreviations used
- Order on recap sheet
- MAR not order sheet
- Pre-checked order boxes

**CONTRAINDICATIONS**

- Dose outside recommended range for age and/or renal function
- Drug contraindicated due to disease: RELATIVE
- Drug contraindicated due to disease: ABSOLUTE
- Drug interaction: RELATIVE
- Drug interaction: ABSOLUTE
- Drug duplication

**INCORRECT INFORMATION**

- Inappropriate indication
- Incorrect/inappropriate route
- Incorrect addressograph

**MISSING INFORMATION**

- Missing dose or strength
- Missing Medication
- PRN with no use parameters
- PRN with no hold parameters
- Missing addressograph
- Missing frequency
- Missing signature
- Missing date of order
- Missing time of order
- Missing pager #
- Missing route

**INTERPRETABILITY**

- "Q" Issues
- Leading/Following Zeros

Comments

Start | Novell-delivered... | GroupWise - Ma... | Microsoft Ac... | Microsoft Word | 2:53 PM

Figure 2. Example of Virginia Mason Satisfaction Survey Layout (data are illustrative only)

Staff Satisfaction Survey Results - 2000				Vice President Report		
Vice President: Sarah Brown				Sections I through VII report the percent of staff members who responded "4" (agree) or "5" (strongly agree) on a 5-point scale. For all items except the Overall item, the scale ranged from 5 = "Strongly Agree" to 1 "Strongly Disagree."		
	VMMC Overall 2000	Sarah Brown 1999	Sarah Brown 2000	Sarah Brown 1999 vs. 2000	! Statistically significant difference * (p < .01)	◆ High response variation
<i>Number of staff responding</i>	3682	567	769			
<b>Section I: Leadership, Internal/External Customer Service</b>						
I would recommend Virginia Mason to friends and family as a good place to work	64.7%	49.2%	64.6%	15.4	!	
The direction in which Virginia Mason leadership is taking the organization is clear	39.0%	39.8%	42.2%	2.4		
I am confident that Virginia Mason leadership is taking us in the right direction <sup>1</sup>	38.8%	NA	43.8%			
I would recommend Virginia Mason to friends and family for their health care needs	77.8%	68.8%	71.9%	3.1		
Virginia Mason provides excellent customer service <sup>1,2</sup>	51.1%	55.2%	51.6%	-3.6		
I feel that all staff members are valued and appreciated at Virginia Mason	34.8%	32.8%	36.3%	3.5		
Virginia Mason does a good job of communication its overall business strategy to staff members	34.5%	41.4%	38.6%	-2.8		
There is a spirit of team work throughout Virginia Mason	40.8%	32.6%	37.9%	5.3	!	
When I call another department, I am greeted warmly and treated like a customer <sup>1,2</sup>	40.3%	37.1%	38.6%	1.5		
I feel that I am part of a team that is working toward providing high quality care	64.4%	63.1%	58.6%	-4.5		
<b>Section II: Relationship With Supervisor</b>						
My direct supervisor is open to hearing staff ideas, concerns, and suggestions <sup>3</sup>	78.3%	69.6%	75.5%	5.9	!	
My direct supervisor gives me enough opportunities to participate in decisions on how my work is to be done <sup>3</sup>	73.8%	64.9%	71.2%	6.3	!	
If I don't have the information I need, I feel comfortable asking my direct supervisor	79.7%	75.6%	76.9%	1.3		
My direct supervisor gives me useful feedback on my job performance on a regular basis	60.6%	50.3%	61.5%	11.2	!	◆

\* Represents a statistically significant difference between the 1999 and 2000 scores for this Vice President

\*\* Statistical significance not calculated for groups less than 25

# Responses not shown for groups less than 6

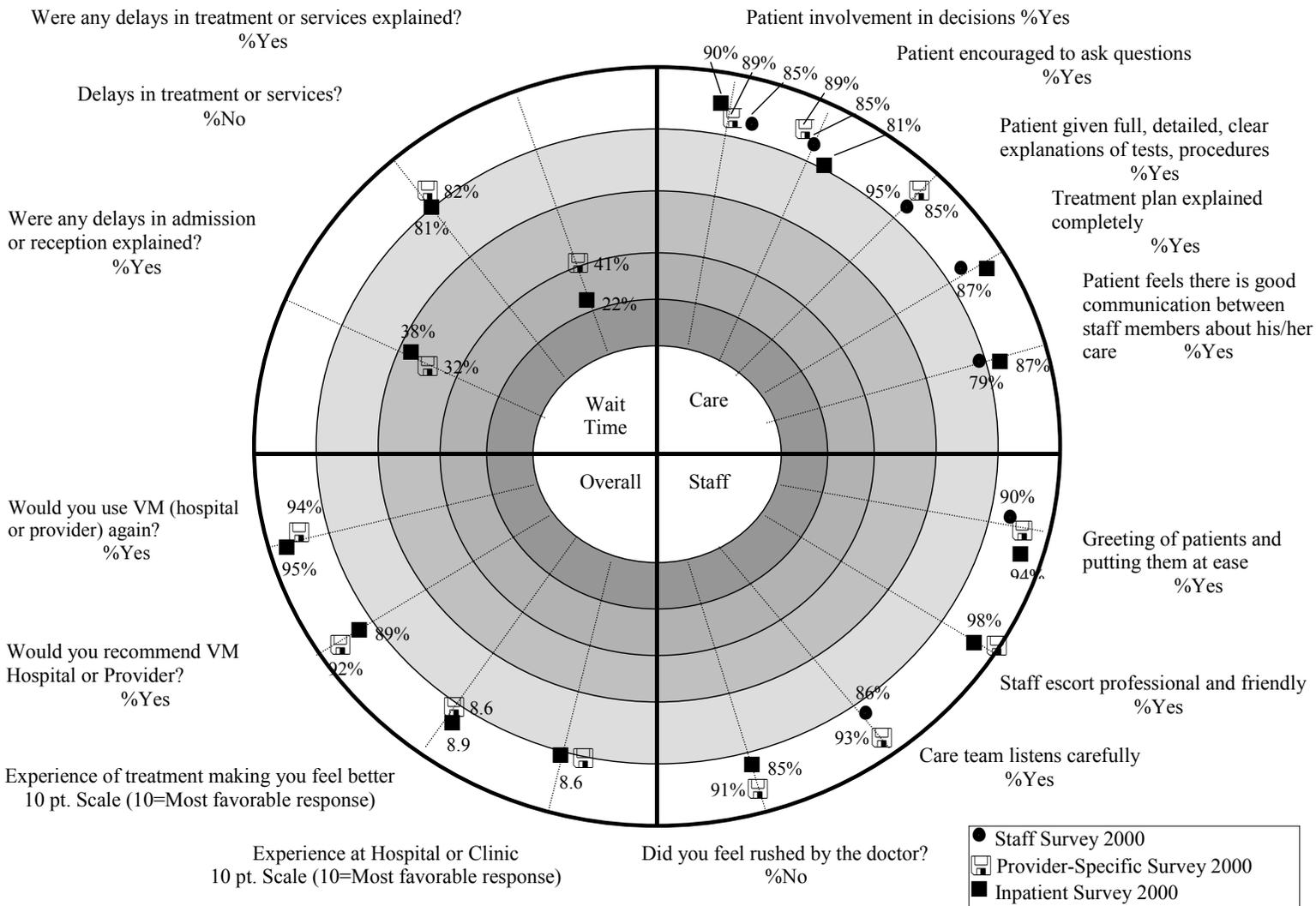
<sup>1</sup> Performance Indicator - VMMC Accountability

<sup>2</sup> Performance Indicator - VP Accountability

<sup>3</sup> Performance Indicator - Manager Accountability

Figure 3. Virginia Mason's 360° View of Satisfaction (data are illustrative only)

**Satisfaction Surveys: Integrated Results**



## References

---

- <sup>1</sup> Nightingale, F. *Notes on Nursing*. (V. Skretkowitz, Ed.). London: Scutari Press, 1992.
- <sup>2</sup> Ulrich, BT. *Leadership and Management According to Florence Nightingale*. Norwalk, CT: Appleton & Lange, 1992.
- <sup>3</sup> Donabedian, A. The end results of health care: Ernest Codman's contribution to quality assessment and beyond. *Milbank Quarterly*, 1989, 67, 233-256.
- <sup>4</sup> Ellwood, PM. Shattuck lecture – Outcomes Management: A technology of patient experience. *New England Journal of Medicine*, 1988, 318, 1549-1556.
- <sup>5</sup> American Medical Group Association. Outcomes management consortia. [http://www.amga.org/AMGA2000/QMR/omc\\_qmr.htm](http://www.amga.org/AMGA2000/QMR/omc_qmr.htm)
- <sup>6</sup> National Committee on Quality Assurance. *Standards for the Accreditation of Managed Care Organizations*. Washington, DC: NCQA, 2000.
- <sup>7</sup> Hurtado, MP, Swift, EK, & Corrigan, JM (Eds.) *Envisioning the National Health Care Quality Report*. Washington, DC: National Academy Press, 2001.
- <sup>8</sup> Burda, D. JCAHO hits a wall with plan on indicators. *Modern Healthcare*, 1994 (March 14), 30-40.
- <sup>9</sup> Joint Commission on Accreditation of Healthcare Organizations. *2001-2002 Comprehensive Accreditation Manual for Health Care Networks*. Oakbrook Terrace, IL: JCAHO, 2001.
- <sup>10</sup> Foundation for Accountability. Measuring what's meaningful. *Accountability Action*. Fall, 1998. <http://www.facct.org>
- <sup>11</sup> Kaplan, RS & Norton, DP. The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 1992, 71-79.
- <sup>12</sup> Nelson, EC, Mohr, JJ, Batalden, PB, & Plume, SK. Improving health care (Part I): The clinical value compass. *Joint Commission Journal of Quality Improvement*, 1996, 22, 243-258.
- <sup>13</sup> Nelson, EC, et al. Report cards or instrument panels: Who needs what? *Joint Commission Journal of Quality Improvement*, 1995, 21, 155-166.
- <sup>14</sup> Gustafson, DH, Arora, NK, Nelson, EC, & Boberg, EW. Increasing understanding of patient needs during and after hospitalization. *Joint Commission Journal of Quality Improvement*, 2001, 27, 81-92.
- <sup>15</sup> Nerenz DR, Zajac BM (Eds.): *Assessing Performance of Integrated Delivery Systems: Analyzing Indicators and Reporting Results*. Volume 5 of Faulkner & Gray's Medical Outcomes and Practice Guideline Library. Faulkner & Gray, NY, 1996.
- <sup>16</sup> Donabedian, A. Evaluating the quality of medical care. *Milbank Quarterly*, 1966, 44, 166-206.
- <sup>17</sup> Chassin, MR. Quality of care: Time to act. *JAMA*, 1991, 266, 3472-3473.
- <sup>18</sup> Kohn, LT, Corrigan, J, & Donaldson, MS (Eds.) *To Err is Human: Building a Safer Health System*. Washington, DC: National Academies Press, 1999.
- <sup>19</sup> HCUPnet, Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD. <http://www.ahrq.gov/hcup/hcupnet.htm>
- <sup>20</sup> Lazar, HL, Fitzgerald, CA, Bao, AT, et al. Early discharge after coronary artery bypass graft surgery: Are patients really going home earlier? *Journal of Thoracic and Cardiovascular Surgery*, 2001, 121, 943-950.

- 
- <sup>21</sup> Kaplan, RS. One cost system isn't enough. *Harvard Business Review*, 1988 (Jan-Feb), 61-66.
- <sup>22</sup> Cleary, PD & Edgman-Levitan, S. Health care quality: Incorporating consumer perspectives. *JAMA*, 1997, 278, 1608-1612.
- <sup>23</sup> Cohen, E. & Goode, TD. Rationale for cultural competence in primary health care. (Policy Brief 1). Washington, DC: National Center for Cultural Competence, 1999.
- <sup>24</sup> MacTaggart, P. Measuring quality of care: Culturally and linguistically appropriate services. (Presentation to AHRQ User Liaison Program – Strategies to Reduce Health Disparities, Tucson, AZ, April 5, 2001).
- <sup>25</sup> Hickey M. (2000). The real issue is culture: Moving from relationship to results. *Frontiers Of Health Services Management*; 16(4): 35-38.
- <sup>26</sup> Ball MJ and Lillis, JC (2000). Health information systems: challenges for the 21<sup>st</sup> century. *Advanced Practice in Acute & Critical Care*; 11(3); 386-89.
- <sup>27</sup> Kovner AR, Elton JJ, Billings J. (2000). Evidence-based management. *Frontiers Of Health Services Management*; 16(4): 3-24.
- <sup>28</sup> Simmons JC (2000). Encouraging a proactive approach to managing diabetes. *The Quality Letter*, 12(2):2-9.
- <sup>29</sup> Wyatt JC (1994). Clinical data systems, part 3: development and evaluation. *Lancet*; 344(8938): 1682-88.
- <sup>30</sup> Godin P, Hubbs R, Woods B, et al. (1999). New paradigms for medical decision support and education: The Stanford health information network for education. *Topics in Health Information Management*, 20(2):1-14.
- <sup>31</sup> Friedman, NM, Gleeson, JM, Kent, MJ, et al. Management of diabetes mellitus in the Lovelace Health Systems' Episodes of Care program. *Effective Clinical Practice*, 1998, 1, 5-11.