White Paper

INGENIX®

Creating Quality Composite Scores:

Challenges and Issues in Physician Quality Measurement

I. Why is quality measurement important?

Concerns about health care quality, safety, and cost are driving the need to measure, compare, and improve physician performance. Two landmark reports from the Institute of Medicine (IOM) identified significant problems in health care quality.[1, 2] One of these reports, "Crossing the Quality Chasm," concluded that there are major gaps between current medical knowledge and actual health care practice. Meanwhile, health care costs have increased at a double-digit rate without the same progression of improved medical outcomes for patients, including quality indicators.[3]

Recent efforts to benchmark quality of care in the United States have revealed substantial gaps in care for different populations. A recent study using survey and medical records data examined 439 quality indicators across 30 acute and chronic conditions to assess the process of care for 6,712 individuals.[4] Only 54.9 percent of recommended care was actually provided. A similar study by RAND revealed that elderly patients with debilitating medical conditions received recommended care only one-third of the time[5] and if the entire health care system performed at or near the top levels, between 37,600 and 81,000 deaths would be avoided per year, saving approximately $3 billion in unnecessary hospitalization expenses alone.[6] A tremendous opportunity exists to improve clinical outcomes, decrease health care costs and have an impact on worker productivity by improving health care quality.

Organizations continue to strive for valid approaches to measure and improve the quality of medical care. Quality measurement at the health plan level has been undertaken through the efforts of NCQA and other national organizations for a number of years using the HEDIS methodology[7]. More recently, the focus has turned to assessing and comparing quality of care at the physician level. This paper discusses the challenges faced by organizations in measuring quality of care and benchmarking physicians across a set of quality measures. The paper addresses four areas:

1. Overview of key methodological issues in measuring quality
2. Three case studies describing how different organizations have addressed these and other issues when measuring quality in their provider populations, including two organizations that use the Ingenix Symmetry Evidence-based Medicine (EBM) Connect tool to assess compliance with quality care
3. Empirical findings using EBM Connect that underscore some of the challenges in measuring physician quality, including obtaining a sufficient number of cases for each physician to support measurement and creating composite measures of quality
4. A summary and discussion

II. Physician quality measurement–methodological and practical issues

Efforts to measure physician quality in providing care to patients have been hindered by a number of methodological and practical challenges, including selecting measures, sample size, and creating composites. This section describes these and other challenges in greater detail.

Identification and scope of relevant performance measures

Quality measurement for a physician should be based on a meaningful and clinically appropriate set of measures with demonstrated effectiveness in the improvement of care which are supported by the literature, guidelines, or a consensus process. Measures should reflect a physician's overall practice or practice at a disease-specific level. Currently, there is no agreement on the specific set of measures that should be included in the calculation of quality for a physician, and how this list should vary by specialty. Health plans and other organizations have sought guidance from national organizations in producing and endorsing "national standards" and these standards have become increasingly available. In many cases, the measures used by an organization to assess quality are a combination of national standards and additional guidelines acceptable to the providers being measured.

Physician specialties that can be measured

A further challenge in selecting performance measures is the limited availability of an adequate number of well-defined, acceptable measures that represent care for some physician

specialty types. The availability of relevant measures and an adequate number of observations per physician will limit the physician specialty types that can be evaluated with a quality performance score. This is particularly true for proceduralists and surgical specialties. The evaluation of alternative quality measurement approaches (e.g., sequencing of procedures, complications after procedures) may be necessary to evaluate the performance of certain specialists.

Minimum number of observations per provider and reliability of measures

Provider performance measurement is dependent on achieving an adequate number of observations for each physician being analyzed. When there are few observations for a particular physician, there is less likelihood that the observations accurately reflect the physician's true patterns of care. The challenge of achieving an adequate number of observations per physician can be highlighted through analysis of the number of observations or "opportunities" available for quality measurement for a typical physician, at the individual measure level, for all measures for a clinical condition, or overall, across all conditions. For example, how many physicians have 20 or more opportunities for colorectal cancer screening? How many have 20 or more opportunities for HbA1c testing for diabetics? How many opportunities are available when evaluating all measures within diabetes or across all conditions for a provider?

Sufficiently large samples are needed to achieve precise measures for a provider or recommended levels of reliability when conducting provider-level analysis – either at the individual quality rule level or where results are aggregated across quality rules, conditions, and patients to create a composite measure. Provider measurement at the health plan level contributes to the challenges in achievin sufficient sample size. Later in this paper, further background on this issue is provided, using results on the number of observations per provider from an application of quality measurement.

Accounting for variability in measurement

A related issue to sufficient sample size is accounting for the inherent variability between individual physicians. When there is significant variation across physicians in the treatment of a particular condition, there is less likelihood that the observed average result for a particular physician accurately reflects that physician's true performance. Individual physician variability results from both known and unknown factors. Known factors, which often are not accounted for, include differences in patient complexity (e.g., unknown biometric test results), patient adherence (e.g., patient refusal of recommended treatments), and local or regional treatment patterns. Both known and unknown factors can be indirectly accounted for by statistical and variance calculations. Credibility adjustment and significance testing increase confidence that the treatment patterns of physicians who are identified as "outliers" are truly outside the expected distribution and not just the result of random variation.

Patient population differences

Risk adjustment takes into account variations in the health status of a physician's patient population. Often these differences are defined by morbidity, comorbidity, age, and gender (case mix adjustment) or a patient's degree of illness (severity adjustment). Many quality measures address patient differences by focusing on a particular condition, or patients of a defined level of severity within a condition. Excluding patients with contraindications from measurement is another approach to remove any bias due to patient differences. However, in some cases, adjusting for patient differences presents a greater challenge. For example, performance measures that identify medication interventions may be particularly sensitive to patient population differences. Patients with multiple medical problems may have more difficulty taking medications that are recommended for the management of certain chronic diseases. Socio-economic status may further impact measured quality. Physicians who care for large groups of uninsured or underinsured patients may face unique quality care challenges (e.g., health care access, inability to pay for care or for medical co-payments and deductibles). These issues should be considered when selecting performance measures, identifying physician comparison groups and benchmarks, or defining quality performance thresholds.

Physician Attribution

Physician attribution is a key step in creating composite quality measures, but it has its own set of unique challenges. There are many methods to attribute physicians to patients

or groups of patients. A first step in attribution is deciding what variables are key indicators of the ties between patients and providers. For example, physician attribution assignment could be based on the highest cost attributed with a particular patient or highest utilization for a given patient in a given year related to each physician. In addition, requirements could include face-to-face interaction with patients or physicians must have an eligible specialty type in at least the last 12 months, if focus on quality is specialty driven. In some instances, when attributing physicians, they may tie based on a certain set of criteria and thus it is important to also set up tiebreaking rules and a hierarchy of tiebreakers to assign patients to physicians. There are many methods associated with physician attribution, and one must weigh the pros and cons of these methods when using these relationships to represent quality measures in health care. The decisions to attribute must be grounded in practical considerations for the market and statistically valid considerations to back up the outputs of decision rules applied.

Creating Quality Composite Scores

A quality composite score is an overall measure of performance for a physician, combining the results of multiple individual quality measures. Creating a valid composite score provides certain advantages in measuring physician quality, including increasing the number of observations available for a physician by aggregating results across rules and allowing a general assessment of quality, overall, or for a particular disease state.

In spite of these advantages, composite measures also present certain challenges. These include:

- Interpretability. While an individual quality rule can describe compliance with a specific quality behavior for a defined condition, composite scores can aggregate performance across a number of quality dimensions and clinical areas. Interpretation and how to improve quality of care related to such measures can present challenges for identifying high and low performers for particular types of conditions and for programs designed for improvement.
- Adjusting for the mix of performance measures. Some quality measures present a greater challenge to physicians in achieving compliance – due to the inherent nature of the processes being measured or the ability of the available data to effectively capture the patients most qualified for the prescribed guideline. Some quality measures also exhibit greater variability across patients and providers. In creating a composite, care should be taken to adjust for the mix of rules, both to account for item difficulty and differences in measure variation. Creating "standardized scores" is one approach to address this challenge.
- Independence of performance measures. A simple composite score approach, which sums numerators and denominators across multiple measures, assumes that the measures are independent and equally important. Yet some performance measures describe a similar pattern or practice attribute within and across conditions. For example, a patient with diabetes mellitus and coronary artery disease (CAD) that has achieved a LDL cholesterol result of less than 100 mg/dL may satisfy two separate LDL result measures – one designed for patients with diabetes and one designed for patients with CAD. Is this acceptable when the same dimension of care, appropriate cholesterol management, is being measured? Perhaps "double counting" in this situation is clinically reasonable. A patient with both diabetes and CAD is at high risk for a cardiovascular event. Therefore, "double counting" in this situation may appropriately reflect the added value of achieving the goal LDL in this high risk patient. "Double counting" and the uncertain independence of performance measures are issues that should be considered when developing a quality composite score method.[8] One possible approach is to combine these measures into a "mini-composite" score before creating a broader composite score. This assumes that measures of uncertain independence can be readily identified.
- Weighing quality measures. All performance measures are not equal with respect to improving clinical outcomes or promoting quality care. Conceptually, unique weights applied to measures or groups of measures could address the differences that certain interventions have on clinical outcomes. For example, higher weights could be assigned to evidence-based measures that are clearly associated with improved outcomes.[8] Measures endorsed by national organizations could also receive higher weight. A further

issue is the weighing of chronic and acute versus preventative measures, such as breast cancer screening and childhood immunizations. For example, preventative measures will typically impact a greater number of individuals and possibly represent a disproportionate amount of a provider's quality opportunities. To what extent do these measures best reflect a provider's quality of care and how should they be weighted?

- Composites or Sub-Composites? A composite quality score can describe a provider's results across all measures, or for a subset of those measures, or a "sub-composite." Potential candidates for sub-composites include aggregating findings across all measures within a condition, such as diabetes. Sub-composites by measure type, such as patient safety, patients receiving appropriate medications, monitoring patients on certain pharmacy agents, and prevention are another option. Aggregating results into broader condition categories, such as chronic, acute, and preventive is another option. As noted above, key considerations in selecting and creating a composite include independence of the measures combined, the weighing of individual rules and sub-composites, and the interpretation and communication of the results to support care improvement.

Balancing efficiency, accuracy, data integrity, and feasibility

A quality measurement tool that uses administrative claims data can quickly assess a large volume of measures for a large number of physicians and patients without requiring new data collection. The process is efficient with respect to resource utilization and cost. Yet, administrative claims data alone may not identify all data or factors that are essential to measuring quality care (e.g., smoking cessation efforts, patient preferences, inability to tolerate specific treatments, previous medication adverse events). Chart review or hybrid methodology may improve data integrity and accuracy but will limit large scale measurement efforts by increasing measurement costs.

III. Symmetry Evidence-Based Medicine Connect (EBM Connect®)

Ingenix Symmetry Evidence-Based Medicine Connect (EBM Connect®) is a health care quality engine that allows organizations to compare the care received by individuals against evidence-based guidelines and other national standards. EBM Connect leverages information from enrollment, administrative medical and pharmacy claims, and lab results to provide a powerful approach to measuring health care quality. To do this, EBM Connect uses three key steps. First, it precisely defines rules for evidence-based practice based on original evidence in peer-reviewed literature, guidelines from medical specialty organizations, and measures developed and endorsed by recognized national organizations. Second, the product identifies those individuals to whom those rules apply. Third, administrative claims data is used to compare care provided to EBM connect rules. EBM Connect identifies patients who have not received recommended services. Patients are assessed regarding their compliance with proven treatments, which is clearly associated with the outcome as well as the cost of care.[9,10,11] The 7.0 version of EBM Connect (released December 2007) offers over 450 quality rules covering more than 30 clinical conditions.

The results from EBM Connect enable users to identify and respond to the gap between clinical evidence and health care practice. Specifically, the findings can be used to:

- Identify opportunities for improvement, including "gaps" in care for individual patients and populations
- Improve physician compliance with prescribed care, such as whether appropriate tests were performed to identify potential adverse events from treatment regimens or whether patients are being screened or routinely monitored as appropriate
- Identify diagnostic tests or treatments that are unnecessary or potentially harmful

- Identify patients with indications of poor disease control such as elevated serum parameters
- Identify patients with low adherence to prescribed medication regimens, information which is not otherwise available to physicians
- Identify patients for referral to case management
- Reduce potentially harmful drug-drug or drug-disease interactions

When run in conjunction with its sister product, Episode Treatment Groups (ETG®), EBM Connect also supports direct comparisons of quality and cost outcomes for a wide array of clinical conditions.

IV. Physician quality measurement–case studies

There is no agreed-upon approach to developing quality scores for physicians. Each organization has unique circumstances which dictate their different approach to developing quality measures and in some cases, rewarding physicians based on how care is provided. The three health care organizations described in this section have addressed many practical and methodological challenges in creating provider quality measures. Two of the organizations used the EBM Connect engine described above to assess quality of care. Each of the organizations is relatively large, together representing more than 35 million members and a large number of physicians. All had similar goals in performing measurement:

- Identify physicians whose clinical practice is consistent with evidence and consensus based standards
- Help health care consumers make more informed choices when selecting physicians
- Support physicians in their practice of better quality care

Although some of these programs included both cost efficiency and quality measures, only the approach to quality measurement is described here.

Case Study 1

This health plan assigns a designation to high performing physicians and encourages patients to use these physicians through lower co-payments and deductibles for their services. Physicians from 17 specialty areas are evaluated using 233 separate measures across 32 conditions from EBM Connect. Clinically relevant, high volume measures are selected within each physician specialty. The attribution of a quality rule to a physician is based on that physician providing the highest total cost of services related to the condition.*

Twenty-four months of claims are reviewed. Each physician must have at least five unique eligible patients and ten total rules in order to be assessed. Each measure is assigned a numeric weight between 1 and 3, depending on endorsement by national organizations and type of quality rule. In particular:

- Twelve measures endorsed by AQA or recognized as IOM priority areas are assigned the highest weight of 3
- The majority of measures, 203, are assigned a weight of 2. Measures with this weight are typically classified as disease management, patient safety, and care pattern measures
- Eighteen measures are assigned the lowest weight of 1. These measures include care pattern of concern and laboratory result measures. Also, some disease management and care pattern measures are assigned the lowest weight of 1

The ratio of compliant measures as ratio of total eligible measures must exceed 70% for the physician to receive high-performance designation. This 70% threshold is used for all physician specialty types. When a physician treats more than one condition, the sum of all eligible rules for the attributed conditions are summarized at the physician level to generate a single overall score. Physicians that receive quality designation are then evaluated for efficiency of care.

Case Study 2

This health care organization also assigns a quality designation to high performing physicians and physician groups and shares results with health care consumers to support informed

*To measure condition-related costs, this organization employs the cost of services assigned to the patient's clinically related episode, as defined by Episode Treatment Groups (ETG); each episode is attributed to only one physician. All condition-related quality rules are then attributed to that physician. Symmetry ETG is a tool used in resource consumption analysis. It identifies each clinical condition that a patient has and assigns each service record and the associated dollars to one of the clinical conditions.

choice when selecting a physician. In some situations, patients are incented to use "quality" physicians through lower co-payments and deductibles. Although an overall quality score is generated for both individual physicians and physician groups, the group score, when available, supersedes the individual score. Two additional dimensions of clinical quality, board certification and NCQA recognition, are also used in the evaluation process but are not reviewed here.

Physicians from more than 20 specialty areas are evaluated using 44 EBM Connect measures, including HEDIS and AQA/NQF endorsed measures. Relevant measures are reviewed and agreed upon by medical staff and executives.

The initial focus of the organization was on quality assessment for specialty providers (e.g., cardiology, endocrinology, pulmonary, obstetrics/gynecology, neurology). Measurement for primary care providers (e.g., internal medicine, family practice, pediatrics) is being phased in at a later date. Each geographic health care market is evaluated separately. For physician attribution, a physician must have at least two face-to-face encounters with the patient within 12 months and must have an eligible specialty type for the condition (e.g., Congestive Heart Failure – Cardiology).

Twenty-four months of claims are reviewed. A quality score is calculated only if a physician or physician group has at least 20 eligible opportunities (i.e., at least 20 "yes or no" outcomes). All measures are weighted equally.

For each measure, a market-based benchmark compliance rate is calculated using all available market measure data for relevant specialists. This compliance rate, which is used for all specialties for which the measure is relevant, is defined as the sum total of measure compliances over the sum total of measure opportunities. The physician/group number of successes is divided by the physician/group expected successes to generate one composite quality index score.

A "z-score" is calculated for each physician/group, accounting for that group's deviation from peers, the inherent variation in their mix of measures, and the number of quality opportunities available for measurement. Groups are ranked by z-score and those in the top 25% within a market receive "quality designation". The quality index for an individual physician is used for informational purposes in contact visits with providers – however, quality designation is based on the group result. This approach allows group opportunity volume and group coordination of care to influence quality designation.

At a later time, approximately 100 EBM Connect measures may be added to the program's quality measurement subset. A weighing system for measures is also under consideration.

Case Study 3

This organization assigns a network designation to high performing, cost efficient physician groups and shares this information with health care consumers to support informed choices when selecting physicians. Depending on the plan sponsor/employer, patients may be encouraged to use "network" physicians through lower co-payments and deductibles.

Physicians are evaluated using eight separate measures. The source of these measures includes Centers for Medicare & Medicaid Services (CMS) and Agency for Healthcare Research and Quality (AHRQ). Measures include recommended medications and screening tests (e.g., beta-blocker after acute myocardial infarction, breast cancer screening), hospital readmissions, and presence of adverse events.

The assessment of quality performance focuses on providers in 12 specialties (e.g., cardiology, gastroenterology, general surgery, orthopedics, and obstetrics/gynecology). The program does not include primary care physicians. Of the eight selected measures, three are used for cardiologists only, three are used for obstetricians/gynecologists only, and two (hospital readmissions and adverse events) are used for all specialties that are evaluated by the program.

For surgeons, physician attribution assignment is determined by identifying the physicians who performed the surgical procedure of interest. For all others, physician attribution is based on highest cost associated with the related ETG episode. Therefore, each surgical procedure and patient episode will be attributed to only one physician.

Twenty-four months of claims are reviewed. A physician or physician group qualifies for calculation of a quality score only if there are 20 or more eligible patient episodes within the 24 months. In addition, an individual physician must have at least 10 quality opportunities for any given measure in order for that measure to be included in the quality score.

For example, a cardiologist may have 50 eligible patients for Measure A and only five eligible patients for Measure B; only Measure A results will be included in the quality score. A measure-specific quality score is calculated based on the percent of total opportunities the measure represents multiplied by the benchmark compliance rate for that measure. For example, a cardiology group with 50% of total opportunities represented by beta-blocker use after acute myocardial infarction where compliance to the measure is 90% will have a beta-blocker score of 0.45.

Measure-specific scores are summed to generate one overall composite quality score. Measures are not otherwise weighted.

Physician groups are rank-ordered by specialty and market. Groups in the lowest 5% quality score ranking are eliminated from network designation. Physician groups in the highest 95% are then evaluated for efficiency. If the efficiency threshold is met, the group receives network designation. Approximately 50% of groups are further eliminated from network designation based on the efficiency score.

In addition to the network designation program, a separate pay-for-performance program generates physician group quality scores. Although this program is not summarized here, the same general methodology is used. The pay-for-performance program uses over 20 measures, including the 8 measures used in the network designation program. The additional measures are AQA endorsed measures.

Table 1: Comparison of Quality Score Methodologies for Three Health Plans

Case Study	Number of Measures	Data Period (mos)	Minimum Observation Threshold	Measure Weighing	Attributing Measures to Providers	Level of Quality Score	Higher Quality Threshold
1	233	24+3	10 (5 pts)	Yes	Highest cost	Overall score only	MD > 70% compliance
2	44	24	20	No	2 or more visits	Overall score only	Group in top 25%
3	8	24	10 (20 ETGs)	No	Highest cost	Overall score only	Group in top 95%

Case Study Summary

Table 1 provides a summary overview of the three organization's quality score methodoligies. As shown in table 1, each organization uses administrative claims data covering 24 months to support measurement. Each also develops an overall quality score for a physician or group and places some emphasis on rules endorsed by national organizations. However, differences do exist in the minimum number of observations required for measurement, physician attribution approach, the weighing of measures, and the threshold used in defining higher quality. The most striking differences have to do with the mix of quality measures included as part of their physician quality monitoring program – ranging from a small set (n=8) to a large set (n=233) of measures. As described above, each organization faced unique circumstances which required a different approach to developing quality measures and in some cases, rewarding physicians based on how care is provided. These circumstances had a great deal to do with the design of the physician quality program and how the results were used.

V. Describing Quality Measurement Opportunities with EBM Connect

EBM Connect results for a large managed care population were used to shed light on some of the challenges in measuring quality of care at the physician level. Clinically relevant measures which were available to evaluate a select group of specialties were assessed and the number of patients observed for each physician based on this set of rules was summarized. The analysis was based on enrollment, administrative medical and pharmacy claims, and lab results data for a population of 1.6 million members in a single metropolitan market. All members were enrolled in the same health plan, which was selected based on the following criteria:

- Complete and consistent claims, enrollment, and lab results information
- Large enough population of enrolled members to support physician-level measures for a sizable number of physicians (e.g., sufficient patients and quality opportunity volume)

- Good network concentration (i.e., reasonable enrollment concentrated in a defined geographic area with potential for a high rate of patients per provider)

Given health plan size and network characteristics, this plan may be more favorable in terms of observed number of quality opportunities available for measurement. However, the results do provide a good qualitative assessment of the different types of issues surrounding quality measurement.

Data was extracted for the 27 months between January 2004 and April 2006 and processed through the Symmetry EBM Connect application. Patients were identified who met the EBM Connect eligibility criteria for a total of 35 clinical conditions (21 chronic conditions, 3 acute episodic conditions, 3 adult preventive screenings, and 8 childhood immunization measures (see Table 2). These 35 conditions were then applied as they relate to physician specialties chosen for this analysis – cardiologists, pulmonologists, endocrinologists, neurologists, family practitioners, internal medicine physicians, and pediatric physicians. The selection of measures and conditions for each specialty to support the analysis was based on clinical appropriateness, the importance of the measure to the specialty, and the extent to which the physician was likely to have significant control of the compliance event. For example, CAD patients prescribed an ACE-inhibitor were selected for cardiology, while the patient's own adherence in filling the prescribed ACE-inhibitor-containing medications over a period of time (minimum compliance 70%) was not selected for this analysis. Table 2 also describes the number of rules available for each specialty, organized by general condition category. As shown in Table 2, more than 140 rules were available for family practice and internal medicine, while endocrinology had the fewest rules available (n=46). A full list of rules and conditions for each specialty included in the analysis is available from Ingenix-Symmetry.

Table 2: Number of EBM Rules Selected for Analysis, by Specialty and Condition Category

Condition	Acute	Chronic	Preventive	Total
Cardiology	2	83	0	85
Endocrinology	0	46	0	46
Family Practice	13	133	9	155
Internal Medicine	11	132	2	145
Neurology	5	55	0	60
Pediatrics	6	51	8	65
Pulmonology	3	45	0	48
Total	40	545	19	604

Patients and clinical conditions were assigned to a physician using two different *attribution methods*.* For preventive measures, the patient primary care provider was first imputed. Since assigned PCP for each member was not available (the health plan employed a non-gatekeeper model), this approach allowed each member with sufficient services to be assigned an imputed PCP. This assignment was made based on that primary care physician with the greatest amount of primary care services provided to a member. For each member, the results for a preventive EBM measure (e.g., breast cancer screening) for the member were then assigned to that provider.

For non-preventive measures (measures related to acute and chronic conditions), the following process was used to attribute a single managing physician, per specialty, to each EBM Connect condition: the EBM "event physician" file was used to count the number of visits and sum the cost of services provided by each physician for each condition, for each member.** The physician with the largest number of visits, per specialty, was attributed to the patient/condition. In case of a tie, the physician with the highest cost was selected. In the end, for each patient and condition, one physician per specialty was assigned responsibility. All of the EBM rules observed for that condition for the patient were then attributed to that physician.

*Since information on network participation was not available for the health plan, both in- and out-of-network physicians were included in the physician attribution methods and in the following analysis.

**The EBM Event Physician file is a standard output from EBM Connect and includes summary details on all of the providers contributing to a patient's care related to an EBM case (clinical condition). This file provides significant value in supporting different approaches to the attribution of quality rules to providers.

Following attribution, the total number of quality opportunities for each rule was summarized. A quality "opportunity" is defined as a unique rule identified for a patient. Patients can qualify for multiple rules. Total opportunities for a physician are the sum of all patient and rule combinations attributed to them.

Completing attribution on our sample, Tables 3a, 3b, and 3c provide a summary of the distribution of the total number of quality opportunities, by specialty, and by chronic, acute, and preventive condition categories. Table 3a describes the number of physicians in our analysis meeting selected opportunity "thresholds" of 5, 10, 20, 30, and 50 opportunities. This summary is reported separately by specialty for acute, chronic, and preventative EBM rules and in total, across all rules. Table 3b describes that same summary, but represents the number of providers that meet each threshold as a percentage of all providers. Finally, Table 3c describes the percent of all opportunities owned by the provider in each specialty and threshold category.

As can be seen in the tables, there is a varied set of quality measurement opportunities on average and there is variation associated with the number of opportunity threshold groups per specialty (5, 10, 20, 30, 50, or more). For **cardiology, endocrinology, and neurology**, only chronic measures are included (in Table 3a) and an average of 97, 244, and 32 opportunities per specialist were found, respectively. It is interesting to note that **endocrinologists** had the highest number of quality opportunities (average of 244) – a result of the larger number of diabetes quality rules available in EBM. In addition, if one compares **neurology** and **pulmonology** to other specialties based on quality measures, there are fewer specialists overall and fewer quality opportunities if one reviews 20, 30, or 50 or more quality opportunities.

Table 3a: Overall Attributution and Opportunity Rate

Specialty	Measure	# of Physicians	Average Opportunities	# with 5 or More	# with 10 or More	# with 20 or More	# with 30 or More	# with 50 or More
Cardiology	Chronic	663	97	566	484	371	331	269
Endocrinology	Chronic	127	244	118	102	78	71	61
Family Practice	Acute	2,231	36	1,318	1,013	757	641	469
Family Practice	Chronic	2,724	111	2,133	1,805	1,424	1,241	1,058
Family Practice	Preventative	3,624	19	1,530	1,211	845	679	451
Family Practice	Total	4,608	98	2,527	2,062	1,637	1,430	1,225
Internal Medicine	Acute	876	18	543	403	230	162	82
Internal Medicine	Chronic	2,016	104	1,630	1,385	1,111	953	794
Internal Medicine	Preventative	2,080	16	915	688	484	336	201
Internal Medicine	Total	2,819	92	1,760	1,466	1,180	1,014	845
Neurology	Chronic	268	32	183	145	107	90	56
Pediatrics	Acute	948	36	556	444	359	308	214
Pediatrics	Chronic	698	12	369	207	110	63	32
Pediatrics	Preventative	753	20	518	382	271	167	82
Pediatrics	Total	1,325	43	762	600	499	433	347
Pulmonology	Acute	34	15	7	2	1	1	1
Pulmonology	Chronic	195	23	133	102	65	49	26
Pulmonology	Total	198	25	134	103	70	49	27

Table 3b summarizes these same findings, presenting the percentage of all physicians in a specialty meeting that threshold. For this comparison, the total number of physicians in the specialty attributed any EBM opportunity is used to compute the percentage. For example, for family practice, a total of 4,608 physicians were attributed one or more acute, chronic, or preventative measures. Of these physicians, 28.6% had 5 or more acute quality opportunities. Cardiologists and endocrinologists represented the highest percentages of specialists with 30 or more opportunities – 49.9% of cardiologists and 55.9% of endocrinologists were observed to have 30 or more quality opportunities.

Table 3b: Percent of Physicians with Levels of Opportunities

Specialty	Measure	# of Physicians	Average Opportunities	# with 5 or More	# with 10 or More	# with 20 or More	# with 30 or More	# with 50 or More
Cardiology	Chronic	663	97	85.4%	73.0%	56.0%	49.9%	40.6%
Endocrinology	Chronic	127	244	92.9%	80.3%	61.4%	55.9%	48.0%
Family Practice	Acute	2,231	36	59.1%	45.4%	33.9%	28.7%	21.0%
Family Practice	Chronic	2,724	111	78.3%	66.3%	52.3%	45.6%	38.8%
Family Practice	Preventative	3,624	19	42.2%	33.4%	23.3%	18.7%	12.4%
Family Practice	Total	4,608	98	54.8%	44.7%	35.5%	31.0%	26.6%
Internal Medicine	Acute	876	18	62.0%	46.0%	26.3%	18.5%	9.4%
Internal Medicine	Chronic	2,016	104	80.9%	68.7%	55.1%	47.3%	39.4%
Internal Medicine	Preventative	2,080	16	44.0%	33.1%	23.3%	16.2%	9.7%
Internal Medicine	Total	2,819	92	62.4%	52.0%	41.9%	36.0%	30.0%
Neurology	Chronic	268	32	68.3%	54.1%	39.9%	33.6%	20.9%
Pediatrics	Acute	948	36	58.6%	46.8%	37.9%	32.5%	22.6%
Pediatrics	Chronic	698	12	52.9%	29.7%	15.8%	9.0%	4.6%
Pediatrics	Preventative	753	20	68.8%	50.7%	36.0%	22.2%	10.9%
Pediatrics	Total	1,325	43	57.5%	45.3%	37.7%	32.7%	26.2%
Pulmonology	Acute	34	15	20.6%	5.9%	2.9%	2.9%	2.9%
Pulmonology	Chronic	195	23	68.2%	52.3%	33.3%	25.1%	13.3%
Pulmonology	Total	198	25	67.7%	52.0%	35.4%	24.7%	13.6%

*Percent computed from number of physicians with one or more "Total" opportunities for the specialty. Note that some specialties only have chronic disease measures and a Total summary is not included.

Table 3c shows the percentage of total opportunities for a measure category owned by physicians in each threshold group (5, 10, 20, 30, or 50 or more opportunities). As shown, even though a relatively small percentage of physicians may have 50 or more opportunities, these providers own a significant share or all patient opportunities for that specialty. For example, only 2.4% of all pediatricians were observed to have 50 or more chronic quality opportunities. However, these providers owned more than a third (35.8%) of all chronic opportunities in the specialty.

Table 3c: Distribution of Percent of Opportunities

Specialty	Measure	Total Opportunities	% with 5 or More	% with 10 or More	% with 20 or More	% with 30 or More	% with 50 or More
Cardiology	Chronic	64,404	99.5%	98.6%	96.0%	94.5%	90.9%
Endocrinology	Chronic	30,979	99.9%	99.5%	98.5%	98.0%	96.8%
Family Practice	Total	449,882	98.4%	97.0%	93.9%	91.4%	86.4%
Family Practice	Acute	80,454	97.3%	94.8%	90.3%	86.9%	78.6%
Family Practice	Chronic	301,058	99.4%	98.6%	96.9%	95.5%	93.1%
Family Practice	Preventative	68,370	95.4%	92.2%	84.8%	79.0%	66.2%
Internal Medicine	Total	257,947	98.5%	96.9%	93.4%	89.9%	84.4%
Internal Medicine	Acute	16,019	94.8%	89.1%	74.1%	64.1%	45.2%
Internal Medicine	Chronic	209,679	99.4%	98.6%	96.8%	95.0%	92.1%
Internal Medicine	Preventative	32,249	94.4%	89.6%	80.8%	69.8%	53.6%
Neurology	Chronic	8,641	97.6%	94.5%	88.6%	83.8%	68.5%
Pediatrics	Total	57,581	96.6%	91.8%	84.7%	76.3%	62.2%
Pediatrics	Acute	34,068	97.6%	95.5%	92.1%	88.4%	77.8%
Pediatrics	Chronic	8,560	90.8%	78.3%	62.7%	49.8%	35.8%
Pediatrics	Preventative	14,953	97.8%	91.1%	80.5%	64.0%	41.7%
Pulmonology	Total	4,939	95.5%	90.8%	79.8%	71.7%	54.9%
Pulmonology	Acute	509	86.8%	80.9%	78.8%	78.8%	78.8%
Pulmonology	Chronic	4,430	96.5%	91.9%	79.9%	70.9%	52.2%

This analysis presented in Tables 3a, 3b, and 3c describes some of the challenges in creating quality scores for providers, by measure type and overall. Even in larger markets such as those used for this analysis, some providers will have less than sufficient experience to support precise summary measures of quality performance. This result is tempered somewhat by the finding that those providers meeting higher volume thresholds are responsible for a significant portion of all quality opportunities in their specialty.

In addition to the type of measure (acute, chronic, and preventive) included in a composite score, there is a varied mix of specific conditions across specialties. At the condition level, quality opportunities have different measurement properties, including frequency and importance as they related to caring for patients. Table 4 shows that across specialty, certain conditions have a significant number of measures and therefore dominate the picture of quality for a particular specialty.

Table 4: Number of Condition Quality Measures by Specialty

Case_Description		Family Practice	Internal Medicine	Neurology	Cardiology	Pediatrics	Pulmonology	Endocrinology
Acute Low Back Pain		5	5	5	0	0	0	0
Acute MI		2	2	0	2	0	0	0
Acute Sinusitis		3	3	0	0	3	3	0
Asthma		6	6	0	3	5	7	0
Atrial Fib		8	8	0	8	0	0	0
Breast CA Scrn		1	1	0	0	0	0	0
Breast Cancer - Part		2	2	0	0	0	0	0
CAD		9	9	0	9	0	4	5
CHF		8	8	0	8	0	8	0
COPD		4	4	0	3	0	5	0
CVA/TIA		7	7	7	6	0	0	0
Childhood Imms		7	0	0	0	7	0	0
Chlamydia Scrn		1	1	0	0	1	0	0
Chronic Renal Failure		11	11	0	8	8	0	0
Depression		6	5	1	0	5	0	0
Diabetes		22	22	11	21	15	11	22
Epilepsy		3	3	16	0	3	0	0
HIV/AIDS		2	2	0	0	0	0	0
Hepatitis C		3	3	0	0	0	0	0
Hyperlipidemia		7	7	3	7	1	0	7
Hypertension		4	4	4	4	4	4	4
Migraine		5	5	5	0	5	0	0
Multiple Sclerosis		5	5	8	0	0	0	0
Obesity		6	6	0	6	1	6	7
Osteoporosis		1	1	0	0	0	0	1
Otitis Media, Acute		2	0	0	0	2	0	0
Pharyngitis		1	1	0	0	1	0	0
Prostate Cancer - Par		2	2	0	0	0	0	0
Rheumatoid Arthritis		8	8	0	0	0	0	0
Sickle Cell Anemia		4	4	0	0	4	0	0
	Acute	13	11	5	2	6	3	0
	Chronic	133	132	55	83	51	45	46
	Preventative	9	2	0	0	8	0	0
		155	145	60	85	65	48	46

*Rules identified as suitable for physician quality measurement for a specialty and included in the analysis described in this paper.

EBM can also support analysis specific to the condition level (see Table 5 for quality opportunities at the condition level for a sample of specialties) to understand the impact of quality measures related to conditions within specialties. In cardiology, for example, the more prevalent conditions of CAD, hypertension, and hyperlipidemia contribute the greater number of opportunities for measurement. In this case, these three conditions represent, on average, 33 or more quality opportunities in our sample, which is much higher than the quality opportunities for other conditions that these cardiologists encounter. The less prevalent cardiovascular conditions (CHF and Atrial Fib) and other conditions observed for patients seen by these cardiologists (obesity, asthma, COPD) provide fewer opportunities. Overall, this table shows that although opportunities for lower volume measures can contribute to an overall composite, they are less likely to support representative summary measures at the condition level for most physicians. This further underlines the importance of a valid approach to creating composite measures of physician quality for specialties. A follow-up to this white paper will fully tackle this problem by presenting a specific approach to creating composite measures across specialties.

Table 5: Number of quality opportunities per condition

Condition	# of doctors	Mean	Count5	Count10	Count20	Count30	Count50	Specialty
Asthma	16	2.0	0	0	0	0	0	cardiology
Atrial Fib	259	6.0	134	53	5	1	0	cardiology
CAD	493	41.4	445	333	253	208	154	cardiology
CHF	233	13.7	231	115	42	19	3	cardiology
COPD	51	1.6	0	0	0	0	0	cardiology
CVA/TIA	76	2.9	14	1	0	0	0	cardiology
Chronic Renal Failure	9	7.2	9	1	0	0	0	cardiology
Diabetes	177	26.2	177	155	92	51	22	cardiology
Hyperlipidemia	428	36.9	328	259	198	160	110	cardiology
Hypertension	528	33.6	361	306	229	185	123	cardiology
Obesity	70	9.8	43	25	8	3	1	cardiology
CAD	14	7.5	9	1	1	0	0	endocrinology
Diabetes	109	189.8	109	92	68	59	54	endocrinology
Hyperlipidemia	95	69.7	74	60	49	44	38	endocrinology
Hypertension	84	38.3	63	51	42	35	23	endocrinology
Obesity	27	12.9	22	14	6	3	0	endocrinology
Osteoporosis	2	1.0	0	0	0	0	0	endocrinology
Acute Sinusitis	858	18.5	540	402	229	161	81	internal medicine
Asthma	668	5.5	265	103	18	4	2	internal medicine
Atrial Fib	172	2.6	21	0	0	0	0	internal medicine
Breast CA Scrn Scrn	1,871	15.8	891	663	453	313	186	internal medicine
Breast Cancer	63	2.5	5	0	0	0	0	internal medicine
CAD	545	11.3	442	214	73	28	6	internal medicine
CHF	196	8.2	192	31	4	2	2	internal medicine
COPD	312	2.0	18	3	1	0	0	internal medicine
CVA/TIA	167	3.1	38	1	0	0	0	internal medicine
Chlamydia Scrn	913	2.9	167	38	7	1	1	internal medicine
Chronic Renal Failure	68	9.5	68	21	7	0	0	internal medicine
Depression	384	2.5	50	7	1	0	0	internal medicine
Diabetes	1,136	52.5	1,136	1,004	715	577	402	internal medicine
Epilepsy	51	2.7	2	1	1	0	0	internal medicine
HIV/AIDS	39	16.2	7	7	6	6	6	internal medicine
Hepatitis C	100	1.5	2	0	0	0	0	internal medicine
Hyperlipidemia	1,339	46.1	1,015	797	612	524	403	internal medicine
Hypertension	1,632	39.1	1,137	932	717	590	434	internal medicine
Migraine	465	4.4	128	42	5	1	0	internal medicine
Multiple Sclerosis	59	4.8	21	1	1	1	1	internal medicine
Obesity	488	11.5	278	167	76	38	14	internal medicine
Osteoporosis	9	1.0	0	0	0	0	0	internal medicine
Pharyngitis	84	1.6	4	1	0	0	0	internal medicine
Prostate Cancer - Par	146	2.8	10	2	0	0	0	internal medicine
Rheumatoid Arthritis	111	3.3	28	3	1	1	0	internal medicine
Sickle Cell Anemia	8	4.0	0	0	0	0	0	internal medicine

VI. Summary

This paper outlined methodological challenges in measuring physician health care quality, provided three relevant case studies for how health care organizations are attempting to tackle these substantive challenges, and finally presented how EBM Connect can be used to describe quality measures across specialties to benchmark physicians. There are a number of methodological issues that must be considered when properly benchmarking quality opportunities in providing care to patients. In addition to those outlined above, a further challenge is the limited availability of data other than administrative claims data in supporting quality measurement information. Health care organizations must continue to move toward incorporating electronic records of lab values, other clinical information from patient records, and finally patient satisfaction to round out a full set of quality opportunities, some of which are even closer to the patient's experience than medical billing information.

A parallel effort that may inform efforts around physician quality measurement is the experience of the CAHPS (Consumer Assessment of Health Plan Surveys: https://www.cahps.ahrq.gov) movement of the late 1990s. This effort addressed a number of the methodological challenges (sample size, reliability, benchmarking, and display of quality data) outlined in this paper when using survey information as the unit of analysis in quality improvement. CAHPS developed standards around composite measurement and reporting frameworks and tested whatever was developed with consumers, patients, and payers. Those studying different approaches to quality measurement using administrative claims as a key source of information could learn from this earlier activity.

In addition to the methodological challenges, we also saw how three sizeable health plans are using programs that identify "high performing" physicians and thus have set up quality measurement systems. The ultimate goal for these plans was two-fold: (1) to help health care consumers make informed choices when selecting physicians and (2) to support physicians in their practice of achieving quality care. An incentive system is sometimes used to encourage customers to see high performing physicians. Although high performing physicians are defined and identified using different methods, these programs share important similarities. Notably, they all use administrative claims data to measure quality and place a high priority on national standard measures. Observation volume, data integrity, and the percent of physicians that can be measured are some of the common challenges of these organizations. Overall, the experiences of these organizations and of others in the industry measuring physician quality illustrate that physician quality measurement is becoming an essential part of measuring and improving health care value. Further, this measurement has been undertaken in the absence of standard approaches to measurement. Topics for future research include the development of best practices around many of the issues raised in this paper, including the development of valid approaches to creating quality composite measures for providers. The next paper in this series will utilize the data set described herein to create quality composite measures and will tackle many of the issues and challenges presented in this paper.

INGENIX®

For Information: 800.765.6034 | ingenuity@ingenix.com
Ingenix, Inc. | 12125 Technology Dr. | Eden Prairie, MN 55344
www.ingenix.com

References

1. Institute of Medicine, Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Academy Press, 2001.
2. Institute of Medicine. Committee on Quality of Health Care in America. *To Err Is Human: Building a Safer Health System.* National Academy Press, 1999.
3. Wennberg JE, McAndrew Cooper M, eds. *The Dartmouth Atlas of Health Care in the United States.* Chicago: American Hospital Publishing, Inc.; 1998.
4. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. *The Quality of Health Care Delivered to Adults in the United States.* N Engl J Med. 2003;348:2635-45.
5. Wenger NS, et al. *The Quality of Medical Care Provided to Vulnerable Community-Dwelling Older Patients.* Ann Int Med 2003;139:740-747.
6. NCQA. *The State of Health Care Quality 2006.* National Committee for Quality Assurance: Washington, D.C.
7. National Committee for Quality Assurance. *The State of Health Care Quality 2004.* Available at: http://www.ncqa.org/communications/SOMC/SOHC2004.pdf. Accessed November 20, 2006.
8. Greene RA, Beckman H, Partridge GH, Thomas JW. Review of the Massachusetts Group Insurance Commission Physician Profiling and Network Tiering Plan. Massachusetts Medical Society, November 2006. (Available at www.massmed.org/GIC_review). Accessed November 30, 2006.
9. Diring JK. *Non-compliance with proven treatments increases healthcare costs. Managed Healthcare Executive*, 2004.
10. Osterberg, BT. *Adherence to medication. N Engl J Med* 2005;353:487-97.
11. McDonald HP, Garg AG, Haynes RB. *Interventions to enhance patient adherence to medication prescriptions. JAMA* 2002; 288:2868-2879.

From North America, call: 1-800-765-6034 • ingenuity@ingenix.com
For a list of Ingenix global office locations, please refer to our website www.ingenix.com

Corporate Headquarters | 12125 Technology Drive, Eden Prairie, MN 55344