USING HEALTH INFORMATION TECHNOLOGY TO TRACK MEDICAL CARE

[Understanding Medical Informatics and Evidenced Based Outcomes Reporting]

Brent A. Metfessel MD

David E. Marcinko MBBS MBA M.Ed

You will never understand bureaucracies until you understand that for bureaucrats procedure is everything and outcomes are nothing.

Thomas Sowell

Computerized information systems are increasingly being used to analyze the cost-effectiveness and quality of care given by physicians. A number of third parties show interest in such information, including health plans, Federal and state governments, and consumer groups. Physicians need clear awareness of the methods used to track their practice patterns, whether the tracking includes the cost of the practice, quality of care (such as frequency of preventive services that a practice provides), or outcomes monitoring. Using information systems for such purposes is part of the growing field of *medical informatics,* which can be defined as the applied science at the junction of the disciplines of medicine, business, and information technology, which supports the healthcare delivery process and promotes measurable improvements in both quality of care and cost-effectiveness (Source: Medical College of Wisconsin). Although a number of definitions of medical informatics exist, this definition is the one most relevant to the application of informatics to the tracking of care processes and physician profiling.

CATEGORIES OF DATA USED TO PROFILE CARE PROCESSES

Having the correct data to support the measures used in practice profiling is key to accurate reporting. The data must be "clean" and as free from errors as possible. Errors in the data may occur due to a number of factors, such as poor diagnosis or procedure coding as well as the miskeying of data fields such as cost values. In addition, the category of data used needs to match the desired measures that one hopes to obtain. For example, if a person or organization wants to look at the effect of a congestive heart failure treatment regimen on exercise tolerance, claims data would not be the appropriate source. Functional status data would need to be collected as well. The following five data categories are of greatest interest in care profiling.

A. Claims, Encounter, and Other Administrative Data: This data category is the most readily available and abundant. Basically all health plans have access to such data, which includes member demographic information such as age and gender, International Classification of Diseases (ICD-9 and ICD-10) codes for diagnoses, Current Procedural Terminology (CPT®) codes for procedures, physician information including medical specialty, NDC codes to identify drugs, cost fields, and other information. Claims databases can become quite large, and for major health plans often include millions of records. Basic quality and cost-efficiency performance measurement are usually done using claims and administrative data. Quality measures obtainable from claims data include frequency of preventive and disease monitoring services (such as the frequency of hemoglobin A1c tests to monitor diabetes), certain complications of care, and proxies for outcomes such as whether a treatment for a chronic condition leads to a decrease in emergency room visits and hospital admissions.

B. Functional Status Data: This category includes subjective data gathered from the patient in terms of his/her view of the illness and the impact of the illness on activities of living, such as whether or not a congestive heart failure patient has the ability to walk up a flight of stairs without significant shortness of breath. The effect of a treatment regimen on such parameters can be performed to determine whether there is improvement in the patient's view, and whether side effects or complications are creating new difficulties for the patient. The SF-36 (QualityMetric, Inc.), a functional status survey with 36 query items, includes a scale that assesses eight health concepts:

- limitations in physical activities because of health problems
- limitations in social activities because of physical or emotional problems
- limitations in usual role activities because of physical health problems
- bodily pain
- general mental health (psychological distress and well-being)
- limitations in usual role activities because of emotional problems
- vitality (energy and fatigue)
- general health perceptions

Functional status data can be an excellent way to measure specific health outcomes in the patient's view, but the limiting factor is that data collection and analysis can be resource-intensive, expensive, and thus functional status data is less abundant than claims and administrative data.

C. Patient Satisfaction Data: This is a subjective measure of what the patient perceives in terms of the level of service quality and care provided by the clinician.

Many managed care organizations (MCOs) consider patient satisfaction an important measure of physician quality. Although not a direct measure of clinical quality, many researchers link patient satisfaction to clinical outcomes. This data, however, is also resource-intensive to collect and requires commitment on the part of the patient to fill out the forms and return them in the mail or on-line. Selection bias may also occur in terms of patient satisfaction data, in that patients who choose to fill out and return the forms may in some cases not be representative of the overall patient population for a physician. More recently, the field has been moving from measuring "satisfaction" to elucidating a more validated and specific "patient experience of care". The Agency for Healthcare Research and Quality (AHRQ) has developed the Consumer Assessment of Healthcare Providers and Systems (CAHPS®), a part of a national initiative to measure, report on, and improve health care quality from the viewpoint of patients and other consumers. Separate surveys are used for evaluating ambulatory care (A-CAHPS) and facility or hospital care (H-CAHPS). In addition, the National CAHPS Benchmarking Database contains over 10 years of CAHPS survey data from commercial and Medicaid plans and is designed to facilitate comparative analysis of individual CAHPS survey results with benchmarks, including national or regional averages.

D. Clinical and Medical Record Data: This category includes lab and radiology results along with other aspects of the medical record. With the advent of government initiatives and mandates to increase the use of electronic medical records (EMRs), numerous companies are now in existence that develop software for EMRs and the competition is intense. Nevertheless, some MCOs and vendors that perform information reporting for MCOs do collect (or intend to collect in the near future) data such as lab

results. Although once again very resource-intensive to collect and analyze, the proper use of such data can provide a much clearer picture of outcomes and treatment progress. Using diabetes as an example, claims data can provide information such as the frequency of performance of hemoglobin A1c tests, but a separate lab result data feed can give the actual value of hemoglobin A1c, and thus the level of control of the diabetes over the past three to four months can be assessed as an outcome measure. In an effort to make lab values more abundant, the recent creation of CPT II codes attempts to create a hybrid between claims data and lab results. For example, HbA1c or LDL results can be coded as to whether a particular lab target or threshold is met (e.g., LDL-C levels < 100 mg/dl) Although not as granular as actual lab results, CPT II codes can significantly increase the abundance of at least basic laboratory data. As claims-based quality measures are refined and enhanced in the future, CPT II codes may increase in importance in an effort to better collect actual outcomes data, and physicians well-versed in the use of the codes could have a distinct advantage in terms of quality performance measurement.

E. Health Risk Assessment (HRA) Data: Although HRA data are not generally used to profile care processes *per se*, such data help to determine which members are at highest risk for chronic illness in the future, such as heart disease. Patients usually fill out such surveys directly, and many Internet sites now include free HRAs and calculation of risk scores. Included in HRA surveys are smoking history, dietary habits, general health questions, level of energy, emotional health, driving habits, and other parameters. Physicians can use such results as guides to ascertain which members need the most intensive intervention and thus help prevent poor future outcomes.

THE PROCESS OF CONVERTING INFORMATION TO VALUE FOR ORGANIZATIONS AND PHYSICIANS

Data streams such as claims inputs are difficult to use in their raw form. Such streams may contain 50 or more fields (data items) per record and thus need further processing in order to be useful in reporting and evaluating physician practice patterns. The following depict the six major steps in converting raw data to usable information that leads to action that benefits organizations (such as MCOs) and physicians (Figure 14.1).

[Insert Figure 14.1]

A. Data Collection: This involves inputting the data into the computer, which may range from a large mainframe to a personal workstation depending on the size of the database. In the past, the data often came in the form of a tape and had to be read into the mainframe computer disk storage. More recently, physicians' offices have been able to directly send the claims data electronically into the MCO system. In this scenario claims data is the data source, along with membership data (containing member demographics and eligibility information), physician identifying information, and pharmacy claims which contain NDC codes, fill dates, fill amount, cost data, and other fields.

B. Data Integration and Mapping: In the managed care industry many different claims systems exist. National managed care organizations and third-party vendors of reporting software generally have to integrate the disparate systems together into a common format. This involves standardizing the data fields so that the same data items appear in the same location in the record and have a specified width. Thus, numeric

fields such as cost will have the same number of digits no matter what claims system the values originally came from, and text fields such as physician specialty will have "mappings" so that the different specialty codes from the various systems that refer to the same specialty will be mapped to a single code in the final database. Such data standardization and mapping is very important to the accurate reporting on physician practice patterns since the input into the reporting programs need to enter the system in a single standard format.

C. Processing of Data Audits: Also known as "data cleansing", the items in the database records are audited to check that they meet basic criteria. Usually the health plan either develops its own software to conduct the checks or purchases the software from a third party. Some examples of basic audits include:

- Age-procedure mismatch, such as pediatric procedures performed on adults.
- Gender-diagnosis or gender-procedure mismatch, such as GYN surgery in males
- Notation of missing data or fields.
- Invalid values, such as an invalid physician specialty.
- ICD-CPT® mismatch, an example being a bunion procedure where the only diagnosis code is asthma.
- Data with out of range values; for instance, a claims record with a cost field value of $1,000,000.

In many cases, the records with errors or audit flags are output as exception reports. In those cases, the health plan would decide whether to keep the record, modify it, or throw it out prior to input into the reporting system.

When a health plan brings a new group of physicians into its network, physicians and their office staff need to keep in mind that the initial mapping of the physician group's data may increase the chance for data errors and thus diligence is needed on the physician side to ensure a smooth, error-free transition to the new database system.

D. Data Grouping for Case-Mix Adjustment: Proper risk adjustment or case-mix adjustment of the data is a necessary component of practice pattern profiles. Such algorithms help to level the playing field among physicians or facilities (such as hospitals) that are being compared. Without such adjustment, a physician that receives a complaint from a health plan that his/her practice is too costly could argue that "my patients are sicker", which may be true in some cases. To adjust for case-mix, the data need to be fed into a grouper that clusters the data into clinical classes or risk groups. The class in which a data record belongs can then be added to the claims record as an additional field(s). These fields are then input into the reporting system along with the rest of the record and used to calculate "expected" or target values for physicians or other comparative groups. The cost of practice is a commonly used value that is adjusted by case-mix, but other metrics can undergo adjustment as well, such as visit rates and procedure utilization.

E. Information Reporting: In this step the case-mix adjusted data is run through the reporting software systems to generate graphical or tabular reports that provide information on practice pattern performance. Such reports may be displayed either in Microsoft Excel or another third party reporting platform. Furthermore, many health plans have developed their own reporting platforms for tables and displays. Often the reports are mailed out to physicians, but an increasing number of health plans have

physician performance reports accessible on the Web, so that physicians and in some cases plan members can readily view the reports on-line. On-line reporting that is accessible to present and prospective patients increases the stakes behind practice performance reporting, since performance results can then directly affect patient traffic for a physician or physician group and thus have economic impact to the practice. Commonly reported items include total cost of practice, cost by service category (such as lab costs, specialist professional costs, and facility costs), visit rates, preventive services rates (such as mammography screening), complication rates, and case-mix adjusted performance ratios (actual/expected cost) or cost variance (actual – expected cost). Many of the numeric measures on a report can undergo case-mix adjustment and the physician given a performance statistic such as a cost-efficiency ratio or a ratio of proportion of quality rules fulfilled for the physician or practice compared to a reference group of peers. A performance ratio of 1.0 means the physician is practicing at the norm for the comparison group, and practice variation can be investigated further through other reports if the performance ratio deviates significantly from that number. Typically, reports are distributed quarterly to physicians and generally cover one to three years of experience.

F. The Value of Information: Both physicians and health plans can benefit from information reporting. Such reporting can open up discussion with peers in practice or with health plan medical directors and result in the wider dissemination of best practices. The consequence of decreased practice variation as a result of methodologically sound reporting improves both the quality and cost-effectiveness of care, since high quality of care can lead to lower cost, especially in the long-term (e.g. savings in emergency room usage, unplanned hospitalizations, and resource use from

complications), although quantifying this effect remains an active area of research. Thus, both the physician and the health plan can benefit.

[Insert Figure 14.2]

CASE-MIX ADJUSTMENT ISSUES IN PRACTICE PROFILING

It is difficult to construct an adequate practice pattern profile without case-mix or risk adjustment. There needs to be an algorithm that adjusts for the mix of clinical conditions encountered in a physician practice. Case-mix adjustment can be made for disease class and in some cases severity within disease class, specialty of practice, benefit plan (such as whether there is a pharmacy benefit), and other features of the data that may affect reported results. Comparing a tertiary care center in New York City to a community hospital outside the city is problematic without adjusted data. The tertiary care center may use more resources, and thus cost more, than the community hospital no matter how exemplary the tertiary care center. In addition, it is difficult to compare a cardiologist to a family practitioner, since in general the cardiologist will see patients of greater severity even within the same illness class. Many MCOs can provide literature or provider relations personnel that can discuss the reporting and case-mix methods as questions arise.

A. Algorithms for Case-Mix Adjustment: A wide variety of methodologies exist that are useful for case-mix, risk, and severity of illness adjustment. A number of third-party vendors, as well as national and government-supported organizations, exist that sell software groupers for case-mix categorization. Since each methodology has different strengths, some MCOs have purchased more than one software package. There

is no such thing as a "perfect" adjuster, since no data system can ever actually "see" a patient to get a complete clinical picture. Existing case-mix adjustment algorithms can be divided into three basic categories.

Clinical quality rule adjusters: An example of a clinical quality rule is "Diabetics who received two or more HbA1c tests annually". Some quality rule software systems have up to several hundred clinical quality rules, derived from clinical care guidelines and from national standards such as the National Quality Forum and the AQA Alliance. Some of these organizations represent collaborative efforts between physician specialty organizations, health plans, and governmental agencies working together for the common goal of quality improvement. Usually, each rule in clinical quality measurement systems is adjusted for clinical condition. For example, there may be separate rules for lipid testing for diabetics and lipid testing for coronary artery disease patients. In addition, quality rules have clinical adjusters for excluding patients with more severe instances or confounding clinical situations. For example, patients who are already blind would be excluded from diabetic retinopathy exam measures, and patients with polycystic ovary syndrome would be excluded from diabetes measure denominators when they were only identified by metformin use.

Episode-based case-mix adjusters: These adjusters are also called condition-based or process-based adjusters. These data groupers typically classify the claims records into episodes of care that track the progress of an acute illness from onset to resolution and include related diagnoses and treatments. Some groupers track illness conditions for specified windows of time whether or not the illness "resolved" within that time. The purpose of these adjusters is to capture the longitudinal process of care for a

single condition. A member or patient can, and often does, have more than one episode of care category or instance during a reporting period. Categories of episodes, roughly paralleling illness classes, may vary from a couple hundred to nearly 1000 distinct categories. There is a trade-off between granularity of classes (more classes can more precisely define illness categories) and cell sizes (more classes may reduce the number of instances in each illness category, leading to less statistical stability in performance measurement)

Population-based or patient-based case-mix adjusters: These adjusters utilize complex algorithms to create risk-based categories based on individual episodes of care, but with a compositing algorithm at the final step that leads to a single illness burden or risk level for a patient. Thus, each patient falls into one category, unlike process-based adjusters where a patient can fit into more than one category. Patients with more complex or major chronic illnesses are given more severe categories than patients with only minor acute illnesses or those who do not access the health care system at all during the period (usually these members are still given a base age-gender risk category). These systems have the advantage of being able to readily identify patients with multiple complex conditions. Typically there are fewer categories with these adjusters, and range from about 20 distinct risk categories to about 200 categories. Some MCOs use population-based adjusters for actuarial purposes, to set case-rates or capitation payments, or for predictive modeling where advanced statistical algorithms, such as multiple regression analysis, are used to predict which patients are likely to use the most health care resources in the future based on their utilization in the past. MCO case-management or disease-management resources can then be targeted to the highest risk

patients in an effort to increase patient quality of life while at the same time decreasing total resource utilization and cost of care, including decreasing emergency room use, hospital admissions, sentinel events, and expensive third-line or fourth-line medications. The ability of a case-mix adjuster to explain variation in resource utilization is determined by the "R-squared" (the square of the correlation coefficient), with the case-mix categories or risk score as the independent variables and a measure of resource use (such as cost) as the dependent variable. Age-gender models have an explanatory power of only about 3 – 7% while publications on proprietary adjusters have generally shown that they explain about 30 – 60% of the variation for retrospective analysis. Prospective explanatory power is somewhat less, usually around 15 – 30%.

B. Calculation of Expected Values: The purpose of a case-mix adjustment algorithm is in the calculation of the expected value of a measure for a physician or facility. The expected value is what a physician "should" obtain based on normative values for the individual case-mix or risk groups. To calculate the value, usually a weighted average is performed where the normative cost, such as a plan average, for each case-mix unit or group is weighted according to the physician's individual experience. Thus, for a physician who saw 50 cases of an expensive disease and 20 cases of an inexpensive disease, the expected value will be much more weighted toward the more resource-intensive illness since more cases were seen (Figure 14.3).

[Insert Figure 14.3]

One limitation of the above expected value calculation methodology is that it is a "parametric" algorithm; that is, it assumes a normal (bell-shaped) distribution of costs.

However, it is well known that medical costs are not normally distributed, but skewed toward the high cost side (positively-skewed or skewed right). This is because the lower bound of costs is at zero cost, but there is theoretically no upper bound. Many systems have outlier exclusion algorithms or methodologies where outliers are capped at a certain cost level; for example, episode or patient costs greater than $25,000 dollars are capped at $25,000, and this value is used for reporting. Other systems may completely exclude outliers, such as excluding the top 5% and lowest 5% of costs from the performance rating calculations. The theory is that outliers are statistical anomalies or catastrophic conditions in which a physician may have less complete control over resource use. The development of more advanced statistical techniques for measurement is an active research area, but runs into difficulty if the techniques become too complex to explain to external stakeholders.

A variety of statistical methodologies have been used to determine whether or not the deviation from the expected value is statistically significant. One method is the use of a 95% confidence interval, which is calculated based on a physician's case-mix adjusted cost. In this method, if the lower bound of the physician's 95% confidence interval is above a cost-efficiency ratio of 1.0, then the physician is considered statistically significant in terms of having a higher than expected cost of practice.

C. Case-Mix Indices: Once an expected value is calculated for a physician or facility, comparison of the physician's actual practice patterns to the expected value(s) can take place. In reporting there exists three basic measures that utilize expected values:

Ratio of actual to expected (actual / expected): This measure is termed a "performance ratio" or an "efficiency ratio". A value of about 1.0 would mean that practice patterns are close to the expected target or plan average. For cost comparisons, a value of slightly below 1.0 might even be more ideal as long as the provision of high-quality care is maintained. For quality of care reporting, a performance ratio can be used to compare the proportion of rules fulfilled (or showing compliance with the measure) as compared to a peer group with a similar rule mix. Since rule compliance is typically a "yes-no" value, different statistics, based on a binomial distribution, are utilized rather than those described for cost of care.

Difference between actual and expected values (actual – expected): This measure is termed the "cost variance" and is very useful for looking at the cost impact of practice variation. An additional advantage of this measure is its approximately normal distribution, unlike performance ratios which are skewed toward the high end. This means that relatively simple statistics can be used to isolate physicians or facilities with high positive cost variances for further analysis. Often, a z-score (number of standard deviations from the mean) of +2 or more is used as the approximate criteria for overly high utilization, although other criteria can also be used. It needs to be noted that a highly negative cost variance can point to care problems as well, in particular problems with patient access to care or underutilization of services, so the reasons for very low cost variances also need to be discovered. The quality score is important in this context as well, as physicians with negative cost variances coupled with low quality scores would point to a high likelihood of underutilization. One property of the cost variance is that the value gives higher weight to more expensive episodes. So physicians that treat more

expensive disease conditions may show higher cost variances even though their performance ratios do not show a strong deviation from unity. Thus, it is recommended that a cost variance value is not reported alone, but that a performance ratio is also taken into account in reporting results.

The ratio of the expected value to the unadjusted plan average (expected / average): This measure is the "illness burden" of the physician practice, or the level of illness in the physician's patient panel. A high illness burden means that the physician or facility treats patients that are more ill than the average physician or facility. A physician with a high illness burden and yet a reasonable performance ratio suggests that the physician is highly effective with complex patients. Health plans could decide to give special attention to such physicians to keep them as active as possible in the network.

D. Other Considerations in Analyzing Case-Mix Methodologies: It is important that physicians understand the basics behind case-mix methodologies, at least for the one or two methods used most for their practice performance reports. Such education may consist of readings provided with the distributed performance reports that explain the algorithm as well as evidence for the algorithm's validity. There are further considerations that are relevant to physicians when dealing with case-mix adjusted reports:

1. Are the reported performance measures adjusted by specialty? The rationale for the additional adjustment comes from the fact that even though a number of specialties may treat congestive heart failure, for example, an internist or family practitioner generally treats less severe cases than would a cardiologist. Thus, even if a report is case-mix adjusted by illness class, the adjuster may not fully account for the

differences in patient acuity within the illness class. Adjusting by specialty will enable a more "apples to apples" comparison and achieve greater physician buy-in to the process. However, for less common illnesses the additional specialty adjustment may cause the cell sizes to become too small, causing the adjustment to lose meaning since there would not be enough patients in some cells for meaningful comparisons.

2. What are the exclusion criteria?: After the case-mix adjustment is performed, it is important that prior to reporting there exists criteria for dealing with outliers. Without such criteria, there is a much greater chance that a good clinician may perform poorly on a performance report since a few high-cost outliers, which may occur due to no fault of the clinician, can strongly skew the case-mix indices and lead to artificially high cost variances and performance ratios. Some methodologies exclude general catastrophic cases, such as members with costs above $25,000, or there may be a calculation where catastrophic members are included in the reporting information but their claim costs are truncated to the criteria amount. This is known in statistical circles as Winsorization. Thus, if a patient has costs of $50,000, the costs will be truncated to $25,000 prior to reporting. This has the advantage of including all patients but the disadvantage of not knowing the actual cost of the patient panel. Another way to exclude outliers involves excluding them at the case-mix class level. This means that illnesses that generally use less resources will have different criteria – in this case a lower high outlier exclusion boundary – than would an illness class that typically has high resource use. As discussed previously, if cost is used as the measure of interest, the distribution curve of cost for a particular illness is skewed to the high side and thus does not look like the bell-shaped normal distribution. This makes developing proper exclusion criteria more complex. For

greater accuracy, a "non-parametric" or "distribution-free" outlier test is useful. One such test was developed in 1993 by Sprent and consists of the following equation:

$$(| X_i - M | / MAD) > Max \qquad \text{(Equation 1)}$$

Where X_i represents any value being evaluated for outlier status, M represents the median (the value for which 50% of sample values are above, 50% below) of the sample (such as all cases in a disease class) and MAD is the median absolute deviation. To calculate the MAD value, first obtain the absolute value of the difference between each value and the sample median. Then, sort the difference scores in ascending order. The median of the difference scores is the MAD value. Max is then the criteria point for excluding outliers. A reasonable value of Max would be 5. Both low and high outliers would be excluded based on this equation.

Outliers still may have useful information in themselves. Consequently, after excluding them from the comparative analysis, it may still be useful to report on them separately since such patients, particularly high outliers, may in some cases be steered to case management protocols.

Other statistical techniques, known as "non-parametric" statistics, are not as sensitive to outliers as are techniques based on mean averages, and thus may not need outlier exclusions. This is because these techniques use the median (50^{th} percentile) rather than mean averages as the measure of central tendency. The use of these methods as they relate to physician performance measurement continues to be an active area of research.

THE CONTENT OF PRACTICE PATTERN PROFILES AND REPORTS

This section deals with the kind of measures and information that a physician may see in a performance report as well as in reports internal to the health plan. Physicians need a strong knowledge base about commonly used metrics in reporting so that the physician can intelligently discuss the report content with his/her peers when needed as well as with the health plan that delivered the reports.

A. Quality Reporting: Typically, good quality care may lead to reduced costs, at least in the long run, since stable patients have fewer unplanned visits, less emergency room usage, and a reduced frequency of hospital admissions, all of which save money.

The *Health Plan Employer Data and Information Set* (HEDIS®) contains measures obtainable from claims, survey, physician, membership, and medical record data. HEDIS® was developed in conjunction with the National Center for Quality Assurance (NCQA) and is a widely accepted specification for quality measures. Consumers, managed care organizations, and accrediting bodies have a high level of interest in the HEDIS® results. The measures included in HEDIS® are updated annually and include the use of preventive services, access to care, level of utilization of key procedures, quality measures for acute and chronic illness care, other physician data such as residency completed, board certification, physician turnover for a health plan, health plan membership statistics, and survey data such as member satisfaction. Although HEDIS® was formerly developed for performance measurement at the health plan level, many measures have also been adapted for use at the physician level.

Most of the measures discussed thus far are process of care measures. However, outcome measures are an important component of quality reporting. There are a number of ways to use data to measure outcomes:

1) **Outcomes obtained from claims data**: Claims data have clear limitations for outcomes analysis, as the data mainly deals with the process of care and does not have information directly pertaining to outcomes except where specified in the ICD-9 codes. However, since the advent of CPT II codes, intermediate outcomes, such as whether or not a lab result is above or below a certain threshold, are now available in claims data. However, CPT II codes are still used relatively infrequently, presenting challenges in their systematic use. Thus, one must rely in many cases on *proxy* measures for outcomes. Proxy measures are process of care metrics that can imply certain outcomes, such as length of an illness episode. The following are some ways to ascertain outcomes of care using claims data:

- Complications of care: The ICD-9 codes directly contain language for denoting outcomes. There exist codes for wound infection and dehiscence, miscarriage in pregnancy, and general surgical complications. The coding of a major infection in a cancer patient on chemotherapy is another example of complications-based outcomes obtainable through claims data.
- Procedure re-performances: Two coronary artery stent procedures within a six month to a year period may imply failure of the first stent. However, a medical record check may ultimately be needed since it could also be a stent placed in a new vessel. Returns to the operating room within a few days of a surgical operation, or an outpatient procedure that turns into an inpatient stay within a few days also implies poor outcomes.
- Readmission rates: Two or more hospitalizations for the same episode of care within 30 to 60 days also imply poor outcomes.

- Episode length analysis: The length that an episode of care lasts can be compared between physicians. Shorter episodes for acute illnesses imply better outcomes unless it is due to the expiration of a patient or poor access to care.
- Medication prescribing patterns: In some conditions the drugs prescribed may imply certain outcomes. A rheumatoid arthritis patient that needs Remicade® probably has a more severe form of the illness. Frequent antibiotic switching for an infectious disease such as pneumonia either implies a resistant organism or difficulties in quality of care.
- Emergency room and hospital utilization: Frequent ER use or hospitalizations for chronic conditions such as asthma or congestive heart failure imply a poor outcome from outpatient treatment.

2) **Outcomes obtained from non-claims data**:

- Patient satisfaction data may be an indicator of outcomes, since patient satisfaction or experience of care (see the previous discussion of the CAHPS® survey initiative) often relates directly to how well a patient has progressed with respect to his/her illness.
- Functional status survey data provide a direct subjective account of the severity of illness or outcome of treatment, depending on when the survey was given. A congestive heart failure patient that reports in a survey that he/she cannot walk up a flight of stairs may show non-responsiveness to treatment that needs addressing.

- Clinical data analysis is becoming important as more and more organizations are adding clinical data to claims, such as lab values. Hemoglobin A1c values, for example, hold the key to how well controlled a diabetic is over the long term.

The difficulty with non-claims data is that collection of such data can be resource-intensive and costly, depending on the sophistication of the information systems available.

B. Economic Reporting: This type of practice pattern profiling emphasizes the economic impact of practice variation. Usually a cost data field is used as the measure of interest, and variation from the norm is often determined through such case-mix indices as discussed previously. Costs may also be broken down into service categories, such as lab, surgical, radiology, professional costs (evaluation and management costs or "E & M"), facility, drug, and other cost categories. Each of these service categories should also be case-mix adjusted so that a performance index and/or cost variance can be provided for each one.

Areas that can be profiled in economic and resource utilization profiling include the following:

- Consulting, specialty, and sub-specialty referral practices.
- Prescription habits, including sample dispensation and using generic equivalents, especially for chronic conditions such as hypertension and Type II diabetes.
- Use of invasive and interventional tests such as angiograms, IVPs, bone scans, and certain biopsies.

- Use of non-invasive procedures and tests such as CT and MRI scans, cardiovascular stress tests, chest X-rays, and ultrasounds.
- Average length of hospital stay (ALOS), surgical operating times, use of assistant surgeons, and other utilization parameters.

If a physician receives a report that points to significant practice variation, the question comes up as to what factor(s) caused the variation. This is where the capability of "drill-down" analysis becomes important. In this method, an area of variation is pinpointed and reports are brought up in greater detail specifically concerning that area of variation. For example, if a physician shows a high cost variance for migraine headache, a drill-down analysis into the disease state may show that the physician uses CT and MRI scans of the head significantly more frequently than his/her peers. The physician can then be educated about the clinically appropriate use of such scans, only reserving them for cases having a high index of suspicion for a tumor. Another method of drill-down analysis, used in a published report by a large independent practice association (IPA) in upstate New York, involved dividing physicians into quintiles depending on adjusted costs, with the lowest quintile (bottom 20%) being least expensive and the highest quintile (top 20%) being the most expensive. This was done for patients with hypertension and separately for ENT patients with non-surgical tonsillitis, adenoiditis, and pharyngitis. Strong differences in prescribing patterns, such as lower use of generic drugs, were found in drill-down analysis between physicians in the highest quintile and lower quintiles for hypertension, and a higher utilization of fiberoptic laryngoscopy for physicians in the higher quartiles for the ENT patients. Methods such as these have the advantage of being

able to quickly get to the likely causes of practice variation and are less influenced by outliers, which may complicate a drill-down analysis. The above algorithm is an example of a non-parametric analysis, since the division into quintiles does not depend on a normal distribution.

C. Drill-down Analysis: If a physician or facility is found to have a significant variance from the norm on a measure, such as cost, drill-down analysis is important to find the reason behind the variance. Episodes of care case-mix adjustment are naturally suited to this kind of analysis, but other population-based groupers also allow drill-down if the clinical categories that are precursors to the assignment of a risk score are used. The idea behind drill-down is to obtain greater and greater detail on an area of interest. Thus, if a physician is found to have a high overall cost variance or performance ratio, a user can select the physician and drill-down into emergency room usage, hospitalization frequency, types of illnesses seen, or procedures performed. Case-mix is useful even for the more detailed reports since if, for example, ER use or the utilization of specified procedures is not adjusted for illness burden the "my patients are sicker" argument can easily hold. However, if the procedures are related to illness classes, physicians can be compared to their peers on procedure use for that illness class.

Scenario: Dr. Jones is a family practitioner who had a high patient load from a single large health plan. These patients under his care had a total of 450 episodes over a two-year period. His case-mix adjusted performance ratio was 2.28 and cost variance was $157,400. Dr. Jones requested a drill-down analysis to determine why his practice patterns showed such a high variance from the norm. One area that the health plan data analysts found had high variance were patients he saw with tendinitis of the lower

extremity. He saw 30 episodes of care for this condition, having a total performance ratio for the illness class of 6.0 and a cost variance of $25,300. On further drill-down, the analyst found that the major cost center included the frequency of MRI scans of the lower extremity for the tendinitis patients. His scan rate was 0.4, which means an average of 4 out of 10 episodes received scans, making a total of 12 scans in all. His peers of the same specialty showed 0.1 scans per episode of tendinitis of the lower extremity. Dr. Jones showed a performance ratio of 3.0 and a cost variance of $10,800. On learning this information, Dr. Jones decided to alter his referral patterns so that his scan rate was brought closer to the norm.

Trending may be useful in physician measurement to look at practice patterns over time. Trending is also a good way to look at response to quality improvement initiatives as well (Figure 1). Although trending as discussed above has value, it presents a number of challenges in terms of implementing in a practical fashion. Few health plans have readily-available long-term data (covering 4 to 5 years or more) required to create usable trend analysis. If episodes of care are used for practice pattern analysis, using short date intervals (such as quarterly intervals) so that there can be more points on the trend chart may lead to statistical instability, since care activity can legitimately vary significantly from quarter to quarter for a single episode of care, particularly for long-term or chronic illnesses. Another alternative would be to include the results from rolling quarterly reports, with each report resulting in a single data point. Including the results of quarterly reports in the trend chart would mean the data lacks independence between points, since each quarterly report is based on rolling date intervals and may include one or two years worth of data. Non-parametric analysis, such as dividing physicians into

quintiles based on cost or quality results as discussed previously, are less influenced by outliers and thus performance on physician measures may be more stable over time.

Dealing With Uncertainty in Practice Measurement

Due to high-profile concerns in terms of the variation in quality of care as well as its affordability, practice pattern measurement is here to stay. Until further advances in this area are created, physicians with unexpectedly poor performance ratios, especially in the area of cost-efficiency, should review their data to determine if there are opportunities to improve as well as potential outlier cases contributing to an aberrant value, as well as looking at the health plan methodology for statistical analysis and outlier exclusions. It is important for the physician or other provider being measured to communicate any issues to health plan personnel where possible. Physicians need to remember that practice pattern analysis is a continually evolving field. Given the state of the art, physicians, specialty societies, and other advocacy groups have a responsibility to work with health plans or other practice measurement agencies to make sure quality improvement is at the forefront, that they are active in giving feedback on health plan practice measurement methods, and that as much as possible a collaborative approach is used in working with health plans and other measurement organizations.

CONCLUSIONS

Health plans, consumers, employer groups, government groups, and accrediting bodies are increasingly asking for more detailed information on physician and facility practice patterns. Important report content issues include the need to balance cost

indicators with quality metrics, use of case-mix adjusters and their properties, drill-down methodologies to help determine the reason for practice pattern variation, and if available trending capability to look at practice pattern changes over time. Where possible, supplementation with member satisfaction data, functional status surveys, and clinical values such as lab results may lead to better information on quality and outcomes of practice patterns. This is a continually evolving field, and we can look forward to further advances in the near future that make quality and cost-efficiency measurement more accurate, with further collaboration between physicians, health plans, and government agencies toward the common goal of quality improvement while working together to increase affordability and access to quality care.

COLLABORATE NOW: Continue discussing this chapter online with the author(s), editor(s) and other readers at: www.BusinessofMedicalPractice.com

Acknowledgements:

The author of this chapter for the first edition was Hope Rachel Hetico; RN, MHA.

References and Further Reading:

Bottle A, and Aylin P. Intelligent information: a national system for monitoring clinical performance. *Health Serv Res* 2008: 43(1 Pt 1), pp. 10-31

Goldfield, Norbert and Boland, Peter: Physician Profiling and Risk Adjustment, Aspen Publishers, Inc., 1996.

Greene RA., Beckman HB, Mahoney T. Beyond the efficiency index: finding a better way to reduce overuse and increase efficiency in physician care. Health Affairs 2008: 20(4): w250-w259.

Hussey PS, de Vries H, Romley J, Wang MC, Chen SS, Shekelle PG, McGlynn EA. A Systematic Review of Health Care Efficiency Measures: Health Serv Res. 2009; 44(3):784-805

Kelley, D. Lynn: How to Use Control Charts for Healthcare, American Society for Quality, 1999.

Marcinko, DE: Healthcare Organizations [Financial Management Strategies]. iMBA Inc, Atlanta, GA 2010.

Metfessel, B., Specialist Profiling Using Claims and Administrative Databases, American Journal of Integrated Healtcare, 1(4) Fall 1998, pp. 168-174.

Patel PH, Siemons D, Shields MC. Proven methods to achieve high payment for performance. J Med Pract Manage. 2007: 23(1), pp. 5-11.

Sachdeva RC, Jain S. Making the case to improve quality and reduce costs in pediatric health care. *Pediatr Clin North Am*. 2009: 56(4), pp. 731-43.

Sheshkin, David J., Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC, 2000.

Vasilevskis EE, Meltzer D, Schnipper J, Kaboli P, Wetterneck T, Gonzales D, Arora V, Zhang J, and Auerbach AD. Quality of care for decompensated heart failure: comparable performance between academic hospitalists and non-hospitalists. J *Gen Intern Med*. 2008: 23(9), pp. 1399-406.

Web sites:

CAHPS®: Assessing Health Care Quality from the Patient's Perspective (program brief):

Agency for Healthcare Research and Quality [2010]

https://www.cahps.ahrq.gov/content/cahpsOverview/07-P016.pdf.