STRATEGIC MANAGEMENT IMPROVEMENT FOR PHYSICIANS AND HEALTH PLANS

[Aligning Incentives Among Stakeholders]

Brent A. Metfessel
David Marcinko

Sometimes when you innovate, you make mistakes. It is best to admit them quickly, and get on with improving your other innovations

Steve Jobs

Physicians and health plans use practice pattern information for a number of initiatives, including network optimization; incentive pool, bonus, or withhold distribution; and physician education. Other stakeholders are also interested in physician cost-effectiveness and quality profiles, including health care consumers, employer groups (both self-insured and health plan customers), accrediting bodies (such as the Joint Commission and DNV Healthcare, Inc) as well as physicians and others involved in day-to-day patient care.

USING INFORMATION FOR DECISIONS THAT AFFECT PHYSICIANS

Health plans, consumers, and employer groups desire to use information from practice pattern profile reports for decision making concerning physicians. There exist several areas where such decisions affect physician practices.

A. Network management

Some MCOs use practice profiles to determine which physicians should be brought into the network, maintained as a network member or given a new level in a tiered network. Some health plans allow from one to three years between when they initiate dissemination of physician profiles and when they use them for network decision-making to educate physicians in the methodology and for practice improvement. A gap between providing initial physician profiles and using them for decision-making may also be provided if there are significant changes or updates to the health plan methodology used for practice pattern profiling, to allow time for education, questions, and feedback.

Practice pattern profiles are not usually used alone in making network status decisions. Other issues such as patient satisfaction results, credentialing, and information on possible sanctions must also be considered.

More recently, some health plans have begun using the Web to display physician practice pattern measurement results directly to health care consumers and employer groups that are customers of the MCO. Some MCOs have used a color-coded or symbolic system (such as a stars or check marks) to communicate physician performance results to employer groups and consumers to help them determine which physicians are best for their needs. Several dimensions may be included in a Web-based report, including quality performance, cost-efficiency of practice, patient satisfaction, and other indicators. Web-based physician reporting is a relatively new paradigm and several more years may be needed to fully ascertain its effect on MCO relationships with physicians, employer groups, and consumers.

Another issue in practice pattern profiling is the relative weighting of quality and cost measurement in physician profiles. MCOs are increasingly adopting the “quality

first" paradigm; that is, first measuring quality of care then only measuring cost-efficiency for those who "pass" quality. The intent of this paradigm is to avoid measuring cost at the expense of quality, and promoting those who practice both high quality and cost-efficient care. In some systems, those who pass quality only will still get a positive quality grade, and those who pass both quality and cost-efficiency will get positive grades in both areas. Patient satisfaction survey results may result in a third grade. Accordingly, an active research area that is just beginning to be explored is the correlation of quality and cost of care. The initial thinking, based on preliminary data analysis, is that high quality of care may increase cost somewhat in the short run (e.g., by prescribing more recommended long-term controller medications for persistent asthmatics, statins for coronary artery disease patients, and increased monitoring for diabetics as recommended in guidelines). In the long run, however, costs would be expected to decrease through fewer "sentinel events", such as emergency room visits and hospitalizations for potentially preventable exacerbations, as well as decreased costs from long-term complications, such as diabetic retinal and kidney disease. To more clearly answer these questions requires large data repositories that cover multiple years of follow-up for each patient in question. Given that many patients do not stay with the same employer or health plan for more than two or three years, initiatives to develop centralized databases across different employers and health plans to track these patients over the long term will be useful for this and other research questions.

Another problem for physicians is that there are typically multiple health plans for which they have reimbursement contracts. The practice pattern evaluation methodologies of each of those health plans can differ. Thus, physicians may feel that they only have

time to concentrate on the profiles from the one or two health plans that have enrolled the largest proportion of their patients. MCOs that use profiles generally have separate categories for those physicians with insufficient data or sample size and thus are not scored. Usually physician performance reporting is performed quarterly or semi-annually based on rolling time periods of one to three years.

B. Bonuses and Incentive Pools

Some MCOs use physician profiles to allocate funds to the top-performing physicians. The MCO may give additional bonuses or preferential allocation of incentive pool funds to physicians that perform well on particular cost-effectiveness and quality indices. Incentive pools are often built based on a certain percentage or "withhold" of dollars that are taken from the physicians' usual reimbursement and placed in a pool. Top performers would be allocated the greatest percentage. One mid-sized health plan in the Southeast paid a 20% bonus to physicians with a case-mix adjusted performance ratio (actual/expected cost) of less than 1.3. More recently, some MCOs are integrating both quality and cost measures together for allocation of incentive or bonus payments. This type of methodology can be a basis of "pay for performance" or "P4P" methodologies where higher performing physicians receive a greater share of withhold or bonus dollars.

C. Physician Education in a Spirit of Partnership

The basic idea would be to develop a continuous quality improvement program that would aid physicians in improving practice patterns in accord with advances in medical care. This method can help promote a collaborative relationship between

physicians and the health plan, but it is also the most resource-intensive to implement on both the physician and MCO end.

WORKING IN PARTNERSHIP WITH HEALTH PLANS

The physician-health plan relationship necessitates a partnership rather than a relationship where the MCO plays a strictly regulatory and payer role. Evidence-based clinical guidelines and technology assessments provide a critical source for scientific grounding in utilization review or in creating coverage policy. A partnership approach leads to greater forward momentum for both health plan and physicians in achieving common goals, although it does require more up-front investment in labor and cost. The components of the health plan-physician partnership relationship include physician education on reporting and case-mix adjustment methodologies, clinical guideline dissemination, medical coverage policy discussion, and education on best practices and care improvement.

A. Physician Education on Reporting and Case-mix Methods

A partnership relationship means that reporting entities, as much as possible, open the "black box" in terms of its case-mix and reporting algorithms and methods. In support of this, the Patient Charter for Physician Performance Measurement, Reporting, and Tiering Programs, supported by health plans, medical organizations, and major employer, labor, and consumer groups has promoted transparency and clear disclosure of measures and methodologies used for physician performance measurement systems, including risk and severity adjustment and statistical standards used.

As for case-mix and risk-adjustment methods, physicians should obtain answers to the following questions:

1. Is it an established methodology using commercial software? Companies exist that are completely dedicated to developing case-mix methods. Less well-known proprietary case-mix adjustment products also exist on the market as well. In addition, some MCOs have their own informatics department that develops a custom case-mix methodology. Such algorithms are not necessarily inferior to the more widely used methods. In fact, custom methodologies developed by smaller vendors or by the health plan itself may show more flexibility in meeting local needs, such as tertiary care practices or urban vs. rural practices. In contrast, the more well-known algorithms have the advantage of being on the market for up to 20 years and thus are more "tried and true".

2. Is it adequately tested? A case-mix package should be tested against a number of scenarios and types of patients, including patients of different age groups, severities, and number and types of co-morbidities. Adjusters should have adequate explanatory power (known statistically as the "R-squared"). This statistic tests the proportion of variation in practice patterns between physicians that are explained by patient-specific factors rather than physician-specific factors. A good patient-level case-mix adjuster should have an explanatory power of about 0.50 or more when year 1 data is used to explain year 1 practice costs, also known as retrospective analysis. When year 1 data is used to explain year 2 costs (prospective analysis), the explanatory power may be 0.30 or less. Retrospective reporting is the mode usually used for practice measurement. Prospective

measurement is typically used for case management or predictive modeling of resource use so targeted interventions can be developed for patients who are expected to have a high illness burden or resource use. Active case management can prevent or delay emergency room visits or hospital admissions, serving both to decrease health care costs and increase quality of life for patients. Further explanatory power improvements may occur when clinical data is added to the claims, such as lab results and electronic medical record data, but at present collection of such data is highly resource-intensive and as yet not widely performed.

Some case-mix adjustment algorithms actually have different models depending on patient demographics, such as a "Medicare model" or a "Medicaid model" that assigns a different risk profile to more elderly patients and to patients on public assistance, respectively.

3. Does the case-mix adjuster make clinical sense? A sound risk-adjustment model should be developed with an abundance of clinician input into the process. Most of the well-known adjusters do have extensive clinical experience built into the packages. Often the companies that create and maintain the models have multi-specialty physician panels that meet periodically and evaluate the algorithm development and enhancement process. Some adjusters are more disease-based, and the categories correspond closely to disease classes, such as "Type II Diabetes". Others have more resource utilization-based classifications such as "Major chronic endocrine" conditions. Others are hybrids between the two, with the higher-level category being a disease class, then a sub-categorization

based on severity or comorbidities that have been shown to influence resource use for the disease class.

5. Is the algorithm an open methodology? Most of the well-known case-mix packages have publications in peer-reviewed journals that describe the basic algorithm used in the adjuster. A health plan's analytic or informatics department usually can provide practitioners with publications or at least publication references. Companies that develop case-mix adjusters also have Web sites that explain key points of their algorithms. For more locally-developed or proprietary case-mix approaches, the developers or provider relations personnel usually can deliver descriptive white papers, mailings, or literature that practitioners can access through the Internet.

6. How are the physician reports developed, and what are the processes behind them? More specific questions to ask in this area include:

- How are the expected or reference values calculated? Do they consist of weighted averages based on risk-adjusted peer norms for each case-mix category? That is, are the norms weighted according to my specific patient experience?
- With what norm am I being compared? Am I being compared to a plan average for each case-mix category or to another benchmark? Is it adjusted for specialty where applicable?

7. Is there drill-down capability? A sound reporting system allows drill-down into more detailed data. Merely an overall performance ratio (e.g. actual/expected cost) is not

adequate for the understanding of how to improve practice patterns. The drill pathways, as much as possible, need to allow the physician to see precise areas of variation that are systematically different from his/her peers or another normative reference value. The cost variance (actual – expected cost) should be significant. Since the cost variance is approximately normally distributed, simple z-scores based on standard deviations can be utilized to select physicians needing further practice pattern investigation. A good rule of thumb is to drill down on physicians that differ from the norm by two standard deviations or more. Note that this includes "underutilizers" as well, since there may exist access difficulties. Such access difficulties may later result in increased emergency room or hospital utilization due to patient illness conditions that worsen due to under-treatment, not to mention the effect on the patients' quality of life. A scatterplot depicting both likely overutilizers and underutilizers is displayed in Figure 1. Furthermore, if feasible, trending of performance indicators is important to determine systematic variation over time, thus underscoring the importance of physicians saving their reports from previous time periods.

B. Clinical Guideline and Technology Assessment Dissemination

Few clinicians can keep up with the full wealth of medical literature in their respective fields. Evidence-based clinical guidelines and technology assessments that summarize medical literature for the diagnosis and treatment of specific medical conditions, as well as the efficacy of various medical or surgical procedures, can be enormously helpful by making the information much more digestible. Individual guidelines are generally structured around single disease classes such as asthma diagnosis

and treatment, or around major medical procedures such as the appropriate use of knee arthroscopy. Most of the over 600 MCOs in the United States utilize some type of guidelines – either evidence-based or an expert opinion consensus – to help analyze and improve practice patterns. This is particularly true of the larger MCOs.

The strongest type of clinical guideline is the evidence-based guideline, which synthesizes the most appropriate clinical studies into a white paper and/or a decision tree diagram. Evidence-based guideline developers generally have staff dedicated to reviewing medical studies. These staff members usually consist of registered nurses, physician scientists, and personnel with advanced public health education. The evidence is graded based on reliability and strength of study design. For such grading analyses, double-blind studies would rate high on the rating scale; case-control studies might rate lower, then review papers, consensus expert opinion, and case studies, which would rate the lowest. The evidence is then combined and synthesized into a clinical guideline that discusses the most appropriate diagnostic tests and treatments for a disease condition, and under what circumstances the relevant procedures should be performed. Most clinical guidelines go through a vetting with a committee of physicians in the specialty area relevant to the guideline, and often include sub-specialty academic physicians. This process acts as a final check and balance to make sure the scientific evidence is interpreted reasonably, including highlighting areas of scientific uncertainty that needs further research prior to making a solid recommendation. Typically, conclusions and recommendations in the guideline are "graded" with respect to the level of certainty of the conclusion.

There are several types of evidence summary documents. A "review paper" is a narrative which summarizes the scientific evidence on diagnosis and treatment of certain medical conditions - or on the efficacy of a medical procedure, but often the criteria for including studies in the review are informal and subjective. This issue is somewhat ameliorated in a "systematic review", where there usually is a rigorous process for identifying studies included in the review as well as grading the strength of the scientific evidence. Where possible, systematic reviews include a "meta-analysis" where statistical techniques are used to combine studies together and pool the samples of patients, creating *de facto* larger sample sizes so that the statistical power is increased. The pool of study patients are then treated as a single study for statistical purposes, and may allow the study results to achieve statistical significance in terms of supporting or not supporting a recommendation, despite the fact that the individual studies in isolation do not show statistical significance. An evidence-based "clinical guideline" combines the rigor of a systematic review for study selection with a consensus process where the conclusions in a guideline are reviewed by local or national physician experts in the area to evaluate the soundness of the recommendations.

Lastly, a "technology assessment" is similar to a clinical guideline in terms of development process and rigor of evidence selection and recommendations but usually centers around the efficacy of particular treatments, tests, or procedures rather than a global medical condition. Usually technology assessments concentrate on high-tech, recently-introduced, or controversial procedures. For example, a clinical guideline may be centered on the topic "obesity management in adults", while a technology assessment may concentrate on a narrower topic such as "surgical treatments for obesity" and would

discuss the relative merits and harms of procedures such as laparoscopic adjustable gastric banding as compared to Roux-en-Y gastric bypass. Technology assessments, either done within the health plan or created by a third party, often form the basis for MCO coverage policies.

The above scenario on obesity management is an example of a "comparative medical effectiveness" paradigm. This goes beyond FDA approval for treatments, which requires one to show safety of a treatment and, in the case of drugs, that the treatment is better than a placebo. In contrast, comparative medical effectiveness looks at how a treatment under question compares with already existent treatments in terms of efficacy and safety, and for which patient subgroups the treatment would most benefit. For example, laparoscopic adjustable gastric banding may provide better outcomes for patients who need less radical weight loss and who desire a device that is removable in case severe side effects develop, unlike Roux-en-Y gastric bypass which alters the actual anatomy of the gastrointestinal tract and is much more difficult to reverse, but usually results in greater weight loss. In another scenario, a new antidepressant may be FDA approved since it shows significant efficacy relative to a placebo, but may not show overall improved efficacy when compared to other antidepressants presently used. However, the new drug may show greater benefit to a subset of patients who show particular symptoms, such as psychosis, assuming the studies are large enough to detect these subgroup differences. This concept may also apply to diagnostic testing as well, where a new test may be compared to a "gold standard", if one exists, in terms of sensitivity (the proportion of actual cases of the condition in question that are correctly identified as having the condition) and specificity (the proportion of those not having the

condition in question that are correctly identified as such). Technology assessments and clinical guidelines offer venues for discussion of comparative medical effectiveness considerations from the scientific evidence that can benefit physician practices.

Disseminating guidelines so that they impact clinician practice is not a trivial task. There needs to exist a clear plan for distributing the guidelines to physicians and creating accountability methods for guideline implementation in clinical practice. One suggested method is to disseminate the guidelines initially to physician leaders who have a strong relationship with both the health plan and local physicians, practice cost-effective and high quality medicine, and have a clear understanding of the practice pattern profiling and reporting process. Physician leaders can take the guidelines and apply them to the needs of local clinicians. In addition, such leaders can significantly help remove some of the physician educational burden from the health plan. Physician leaders can educate other physicians through didactic lectures, discussion groups, and one-on-one meetings.

The Institute for Clinical Systems Integration (ICSI) is a strong proponent of the value of evidence-based clinical guidelines, and cites the following objections that make their implementation and acceptance more difficult. These issues generally apply to technology assessments as well:

- **Guidelines are a legal hazard**: There is a fear that following a guideline that turns out to be wrong increases the risk of litigation. Good guidelines, however, are evidence-based and not opinion-based drivers of care. Furthermore, once a review of the literature takes place and is synthesized into a preliminary guideline, multi-specialty physician focus groups review the guidelines prior to finalization. The strength of evidence supporting each conclusion is usually stated,

highlighting areas of remaining scientific uncertainty. "Evidence hierarchies" are often used as aids to grading recommendations, with meta-analysis, systematic reviews, and randomized controlled trials being at or near the top of the hierarchy in strength, with narrative reviews, case reports, and medical opinion pieces being considered the weakest forms of evidence. This provides additional checks and balances to guideline development.

- **Guidelines are cookbook medicine**: Guidelines are just that – guidelines. Each patient should be provided treatment according to his/her individual needs. Evidence-based clinical guidelines are based on extensive reviews of the literature and are applicable to the vast majority of cases for a particular clinical condition but not necessarily all cases. In the case of practice pattern evaluation or profiling, comparisons of such patterns to medical guidelines can help identify overall *systematic* variations from the norm rather than variations due to particular patients with special needs.
- **Guidelines do not work**: When used as the sole basis for practice improvement, this statement contains some truth. However, when incorporated into a systematic continuous quality improvement approach, they have been shown to improve practice patterns and reduce variation.
- **Physicians will not use guidelines**: Once physicians know that the guidelines are based on a sound review of the medical literature, practitioner buy-in greatly increases. In addition, clinicians need to realize that clinical guidelines are only one part of the total treatment picture since a team approach to patient care is becoming the norm.

- **Guidelines need validation through actual outcomes data**: This is correct when based on a continuous quality improvement approach, but is incorrect if outcomes are based on individual events. Local implementation of guidelines can be compared to outcomes data one or two years after implementation. Depending on the actual level of practice pattern improvement, minor alterations can be made to the guidelines to reflect local needs.

National guidelines in some cases may need adaptation to local patient needs and concerns. For example, a practice in a major metropolitan area where specialty care is readily available differs in major ways from a rural practice which is based more on primary care. Practices where many patients are poor or on public assistance also differs from practices in affluent areas. When used as basic guides to appropriate practice, however, clinical guidelines can significantly decrease practice variation.

With the recent emphasis on evidence-based medicine and on decreasing the time lag between evidence publication and its effect on actual patient care, a number of agencies have added clinical guideline and technology assessment development to their task lists. Such agencies include specialty societies such as the American College of Cardiology (ACC), private companies and non-profit organizations, governmental bodies such as the Agency for Health Care Research and Quality (AHRQ), and MCOs that review the scientific evidence for the purpose of determining coverage policy. MCOs may post medical coverage policies on the Web for physicians to access, and these generally contain narrative justifications (often with evidence grading) in terms of why a particular procedure or diagnostic test may or may not be covered based on level of

efficacy shown in scientific studies. It is important to note that for many high-tech or new procedures, different MCOs may have somewhat different coverage policies based on variation in terms of interpreting the evidence, especially in areas where the science is less certain.

Even among non-MCO affiliated guideline developers, recommendations may vary especially where some uncertainty exists in the scientific studies or the results of studies conflict. One case-in-point is new guidelines by the U.S. Preventive Services Task Force released in November 2009 which decreased the frequency of mammograms from annually to once every two years, and changed the recommended threshold age for screening from 40 to 50 years old, although the guideline does state that screening before 50 years old should be an individual decision taking patient factors and values into account. Much of the impetus for this change is due to fears of over-diagnosis and potential patient exposure to unnecessary treatments. However, the American Cancer Society still maintained the recommendation for annual screening starting at age 40. Variation in recommendations between guidelines is not unexpected given the difficulty of designing rigorous studies, such as randomized controlled trials with long-term follow-up, as these studies can be resource-intensive and funding is often limited. Those that find adequate funding may be funded by drug companies or other organizations with a financial interest in the reported results of the study. Although this alone does not necessarily invalidate the study's conclusions, the study still needs to be evaluated with a critical eye.

The development of accountability programs that track how physicians follow basic guideline standards is critical to continuous quality improvement programs. The

program needs an approach that enables physicians to positively view the accountability process rather than consider it a threat or an attempt to take away clinical discretion from the physicians. Information systems exist that measure compliance to simple guidelines, such as providing beta-blockers after an acute MI, prescribing ACE inhibitors for CHF, and the presence in the claims record of a follow-up visit 30 days after a psychiatric hospitalization. Creating more sophisticated measures that reflect the true richness of clinical guidelines is also of great importance. Many of these measures can be obtained using commonly available claims and administrative data. Categories of measures for guideline compliance include the following:

- A procedure or treatment is commonly indicated for a clinical condition and should generally be performed. Many of these measures have already been developed, such as the measurement of hemoglobin A1c testing frequency for diabetic members. There is a difference of opinion in the field as to whether measures should be "all or nothing" or whether "partial credit" may apply. Some have advocated that clusters of measures for common conditions need to be performed in their entirety to receive credit. For example, diabetic measures include annual lipid screening, twice-yearly HbA1c testing, nephropathy testing, and annual retinal exams. In this paradigm, all four measures need to be fulfilled to receive credit. Others advocate for four separate measures, allowing credit for whatever measures show compliance. Most systems at present support the latter of the two paradigms. Another method is to use "approximate reasoning". For example, if a diabetic member receives the HbA1c test just once per year rather than the recommended two tests yearly, any "guideline compliance score" would

be “partial credit”; that is, the score would not be as great as if three tests were given but more than if the member had no tests in a given year. This “approximate reasoning” would also apply to the other suggested guideline compliance measure categories that follow, although there may be some practical limitations on implementing this more complex approach.

- A procedure is indicated, but only after a certain time interval from the illness onset. For example, a new episode of low back pain is usually treated conservatively for about 30 days. If the pain persists after the time period has elapsed, an MRI should be performed. However, multiple MRI scans for the same episode of care for back pain would be discouraged.
- A procedure is indicated but only after another procedure is performed first, such as the need in certain conditions to perform a screening lab test prior to a more extensive diagnostic workup.
- Issues concerning inpatient utilization and setting of care: This includes hospital admissions to perform surgeries generally done on an outpatient basis, consistently long lengths of stay for various illnesses (exposing patients to risks such as nosocomial infection), and unnecessary use of assistant surgeons.
- Pharmaceutical practice patterns: These measures range from simple metrics such as the use of beta-blockers after acute MI to the appropriate use of first and second line hypertension medications prior to third line medications in patients with new onset hypertension.

In the past several years, medical care processes and procedures with the strongest level of recommendation and consensus have found their way into physician quality measurement algorithms. National agencies such as the National Quality Foundation (NQF), and the AQA Alliance (formerly known as the Ambulatory Care Quality Alliance, a consortium of the American Academy of Family Physicians, American College of Physicians, America's Health Insurance Plans, and the Agency for Healthcare Quality and Research) have recommended measures for such use. The AQA Alliance is significant since different stakeholders, including health plans, government agencies, and physician societies, are able to work together for the common goal of quality improvement. For MCOs or other organizations that have quality measurement programs, claims data are used to capture diagnoses given and procedures performed in the context of these diagnoses. This is matched with the algorithm to determine if there was compliance to the clinical recommendation (or "rule") in question. A composite score (usually a percentage of rule "firings" or opportunities with a compliance score of "yes") can be displayed for each physician and compared to either a threshold expected percent compliance or compared to a reference or to a peer group and statistically tested to discover if the physician varied significantly from the reference or peer group in terms of proportion of rules compliant.

One limitation to physician quality profiling is that most rules are limited to process measures. That is, whether a patient filled a recommended prescription or an appropriate diagnostic or monitoring test was given. Although these process measures have been shown to correlate well with future outcomes, actual outcomes, other than coded procedural complications and related items, are more difficult to measure. If a

peer-based reference group is used for comparison, and the size of the patient panel or number of rule firings are high enough for a physician, some factors affecting all physicians, such as patient compliance, will also be present in the reference group and thus "wash out" between the reference group and the physician. Recently, CPT II codes have been introduced to the Current Procedural Terminology (CPT®) codes by the American Medical Association (AMA). These codes apply to some intermediate outcome data not previously ascertainable with claims data, such as whether or not a patient's LDL level was below 100 mg/dl (recommended level for diabetics), between 100 and 129 mg/dl, or 130 mg/dl or greater (a relatively poor result, likely necessitating more aggressive treatment or encouragement of patient compliance). New codes are also in existence that can designate patient refusal or non-compliance with treatment, which may aid physicians in select cases in terms of performance measurement.

Physician practice profiles can be further improved through the use of electronic medical record (EMR) data in the future, hopefully providing further insight into actual patient outcomes. The use of EMR data in physician performance measurement would be a significant advance. However, obtaining this data can be expensive and natural language processing computer algorithms are not yet mature. In the meantime, claims and administrative data, which are plentiful and inexpensive, will continue to be the primary data source for physician practice measurement. This underscores the importance of accurate coding of diagnoses and procedures in the physician's office, to make sure physician measurement accuracy is optimized within the present limitations of the field.

DISEASE MANAGEMENT

One area where technology assessments, clinical guidelines, and EMR data can make a true difference in patient care is in disease management. The Disease Management Association of America (DMAA) defines disease management as "a system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant". Disease management supports the physician-patient relationship and places particular significance on the prevention of exacerbations and complications of chronic diseases using evidence-based clinical guidelines and integrating those recommendations into initiatives to empower patients to be active partners with their physicians in managing their conditions. Typically, targets for disease management efforts include chronic conditions such as asthma, diabetes, chronic obstructive pulmonary disease, coronary artery disease, and heart failure, where patients can be active in self-care and where appropriate lifestyle changes can have a significant favorable impact on illness progression. The DMAA also emphasizes the importance of process and outcomes measurement and evaluation, along with using the data to influence management of the condition. Although claims and administrative data can be used to measure and evaluate selected processes and outcomes, EMRs will be needed to capture the full spectrum of data for analyzing illness response to disease management programs and to support necessary changes in care plans to improve both intermediate outcomes (such as lab values), and long-range goals (such as the prevention of illness exacerbations, managing comorbidities, and halting the progression of complications).

PHYSICIAN EDUCATION

Physician leaders can once again have high influence on physician practice patterns as long as they are trusted colleagues of local physicians. Such physician leaders should develop strong relationships with medical directors of dominant health plans in the geographic area. The medical directors can then educate physician leaders on plan-wide problems and issues who can then educate local physicians and other care providers. These leaders can also have one-on-one sessions with physicians having significant practice variation in an area, showing them peer-based comparisons and allowing feedback from the physicians. This discussion can then lead to the dissemination of guidelines and best practices to aid the physicians in improving their practice patterns. Other methods of education include discussion groups, written material, and didactic lectures and conferences. These latter methods work best when educating multiple physicians on overall or high-level practice pattern issues. Physicians and other clinicians are busy people; therefore, education must be conducted efficiently. Furthermore, physicians should be encouraged to discuss key performance reports with other physicians who receive similar reports. These discussions will help institute what is called the "Hawthorne Effect", which means that merely having the knowledge of practice pattern variation will enable physicians on their own to seek ways to decrease their variation from peer practices. Physicians do not like to be told how to practice medicine. However, it is ingrained in their culture to care about how they compare to their peers. This desire can be healthy and result in enhanced practices.

Given the intensity and high workload of most physicians, a typical practitioner who has patients from multiple health plans may only be able to concentrate on the one or

two health plans that have the most patients in a physician's practice. Those involved in quality improvement and physician education should keep this in mind and maximize the impact of reporting.

GOALS OF PERFORMANCE IMPROVEMENT

The major goals of performance improvement are twofold: First, for a particular practice pattern measure, the desire is to narrow the practice variation around present health care norms. For instance, the spread of the distribution among physicians of a cost variance measure should decrease with process improvement. Second, clinical guideline-based "best practices" can be utilized to move the entire physician population toward better cost-efficiency and quality. Although best practices may be guideline-based, they should be adapted to local considerations and evaluated periodically through actual outcomes analysis. Such outcomes measures may include:

- Cost-efficiency improvement, showing a decrease in unnecessary resource utilization.
- An increase in the performance of preventive measures, such as childhood immunizations and various screening tests such as breast and cervical cancer screening. This may increase costs initially but will likely more than pay for itself through a decreased illness burden and cost in the future.
- A decrease in episode length for acute illnesses, usually implying a quicker resolution of symptoms.
- A decrease in emergency room visits and unplanned hospital admissions.

- A decrease in the rate of “sentinel events” such as status asthmaticus, hemorrhage during pregnancy, diabetic ketoacidosis, and ruptured appendix.

Most of these measures can be obtained using commonly available claims and administrative databases, although future supplementation with clinical and functional status data will only increase the reliability and scope of outcomes analysis.

ASSESSMENT

Today, many professional and specialty societies, initially wary of physician practice measurement, have begun to embrace the practice if used appropriately. They realize that health care costs are becoming more unaffordable and that evidence-based medicine is often not adhered to and care quality variation is high. For example, The American Academy of Family Physicians (AAFP) recently created documentation on the appropriate use of physician measurement, including making sure that the reporting and practice pattern evaluation is done with quality improvement as a prominent aim and to ensure that cost of care measurement is not the sole metric provided, that severity adjustment is applied, that the physician-patient relationship and sufficient access to care are supported, and that the limitations inherent in measurement practices - including scientific uncertainty, use of claims or administrative data to approximate actual clinical practice, and statistical testing algorithms – are acknowledged and the methods of measurement are transparent and open to feedback from medical practitioners.

CONCLUSION

Health plans use physician practice pattern information for many reasons. Regardless, it is imperative that the relationships between physicians and MCOs are in a spirit of partnership. Understanding the processes described in this chapter will assist in establishing a productive relationship among all stakeholders, including patients, hospitals, health plans, accrediting bodies, and other physicians for the development of their referral base and enhancement of career satisfaction.

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:

Bergman DA. Evidence-based guidelines and critical pathways for quality improvement. *Pediatrics* 1999: 103 (1 Suppl E) piii. 225-32

Cutler TW, Palmieri J, Khalsa M, and Stebbins M. Evaluation of the relationship between a chronic disease care management program and California pay-for-performance diabetes care cholesterol measures in one medical group. *J Manag Care Pharm* 2007: 13(7), pp. 578-88.

Curtin K, Beckman H, Pankow G, Milillo Y and Greene RA: Return on investment in pay for performance: a diabetes case study. *J Healthc Manag* 2006: 51(6), pp. 365-74.

Delon S and Mackinnon B. Alberta Health CDM Advisory Committee: Alberta's systems approach to chronic disease management and prevention utilizing the expanded chronic care model. *Healthc Q* 2009: 13, Spec No: 98-104.

Frankel HL, Fitzpatrick MK, Gaskell S, and Hoff WS: Strategies to improve compliance with evidence-based clinical management guidelines. *J Am Coll Surg* 1999: 189(6), pp. 533-8.

Grube ME and Kaufman K. Positioning your organization for success in the new era. *Health Finance Manage* 2010: 64(1), pp. 54-62

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

Metfessel, B: An automated tool for an analysis of compliance to evidence-based clinical guidelines. *MEDINFO 2001*, V. Patel et al. (Eds) © 2001 (International Medical Informatics Association)

Milstein A, Lee TH. Comparing Physicians on Efficiency; NEJM 2007, 357(26): 2649-52.

Nolan T., Berwick DM. All-or-none measurement raises the bar on performance. JAMA 2006; 295(10):1168-70.

Onion CW and Bartzokas CA: Changing attitudes to infection management in primary care: a controlled trial of active versus passive guideline implementation strategies. *Fam Pract* 1998: 15(2) pp. 99-104.

Sandy LG, Rattray M., Thomas JW. Episode-based physician profiling: A guide to the perplexing. J Gen Intern Med 2008; 23(9): 1521-4

Web sites:

American Academy of Family Physicians: Tiered and Select Physician Networks. http://www.aafp.org/online/en/home/policy/policies/t/tieredselphynetworks.html. Accessed on December 15, 2009

American Academy of Family Physicians: Physician Profiling, Guiding Principles. http://www.aafp.org/online/en/home/policy/policies/p/physicianprofiling.html. Accessed on December 15, 2009

Consumer-Purchaser Disclosure Project: Patient Charter for Physician Performance Measurement, Reporting, and Tiering Programs: Ensuring Transparency, Fairness, and Independent Review (April 1, 2008). http://healthcaredisclosure.org/docs/files/PatientCharter.pdf. Accessed on January 7, 2010

Disease Management Association of America (DMAA): Population Health. DMAA Definition of Disease Management; http://www.dmaa.org/dm_definition.asp Accessed on January 13, 2010

THE END