

**Massachusetts Health Plans' Use of Selected Quality
and Utilization Management Tools**

John D. Freedman, M.D., M.B.A.

Bruce E. Landon, M.D., M.B.A.

Executive Summary

Introduction

This report presents results from the Massachusetts Medical Society's (MMS) first report on selected medical management practices by Massachusetts health plans. In particular, the report focuses on how health plans in the state have implemented prior authorization programs for high cost radiology and a selected drug class, pay for performance for physicians, and tiering for both primary care physicians and specialists. The report has two major goals: to describe the implementation of these program by Massachusetts health plans; and to determine how well the programs and policies put in place by the health plans conform with guidelines issued by the MMS in these areas. This report presents a description of the elements of each of the health plan programs and an assessment of the extent to which each element conforms to the MMS guidelines based on a review of available public information and interviews with health plan representatives conducted by the authors.

Prior Authorization Programs

Utilization management for high cost radiology has become widespread in Massachusetts. Consequently, most differences among health plans relate to details of the implementation of their individual programs. Most of the programs are run by a contracted vendor and, in some cases, the same vendor works with several of the health plans. Among the large health plans in the state, only one has not instituted a broad-based program. In addition, four health plans have

limited their program to prior “notification” wherein the ultimate decision is non-binding and left in the hands of the members’ physicians. In most cases, the programs are implemented so as to conform to the MMS guidelines. For instance, all of the programs exempt emergency studies and most of them have alternative arrangements available for urgent studies, including retroactive notification up to two days after a study has been performed for several of the programs. Similarly, convenient access is available via telephone or the internet.

Nonetheless, all of these programs place an additional administrative burden on practicing physicians for which no compensation is received. In addition, the MMS criteria espouse the principle that physicians who conform to guidelines should be exempted from the PA program. The few plans who allow exemption from their PA program do so narrowly and generally still require notification for each procedure. Finally, whereas the MMS guidelines specify that reviews should be performed by a practicing Massachusetts physician who regularly orders the test in question, denials may be issued by non-practicing physicians from any specialty.

Pharmaceutical prior authorization programs are generally driven by the availability of less expensive alternative treatment options or an attempt to prevent the use of a drug for a non-covered condition, rather than trends in utilization. All of the health plans have a Pharmacy and Therapeutics Committee that advises the health plan on formulary issues. Seven of the nine health plans studied require some form of prior authorization for anti-fungal drugs, the category of drugs we examined in this report, mostly because this class of drugs

is used commonly to treat an uncovered cosmetic condition. The health plans generally rely on submission by fax, and turnaround is two days or less. Similar to the radiology programs, denials may be issued by non-practicing practicing physicians of any specialty or, in many cases, pharmacists.

Pay for Performance

All of the major local health plans have instituted pay for performance (P4P) programs in Massachusetts. In contrast, although the three national plans have implemented programs in other states, they have not yet done so in Massachusetts. The most commonly used P4P programs in Massachusetts generally are aimed at the group or contracting unit and focus on primary care. Many of the programs are customized. Because of this, most feedback of results is through local practice leaders, and physicians may not receive direct individualized feedback on their performance.

The MMS guidelines support the use of P4P to achieve improvements in quality. In order to do so, physicians must receive timely and actionable data, which some plans provide. In addition, the incentives can be designed so as to promote the adoption of tools to facilitate population health management (e.g., electronic health records). In general, most of the health plans share goals and data on performance either prior to or early in the measurement year. As noted, however, these data are frequently not at the level of the individual physician. In several cases, these data are updated throughout the year. Several of the health

plans include explicit measures of technology adoption, and others have separate programs aimed at infrastructure improvement to support P4P activities.

Physician Network Tiering

Most tiering activities in the state were implemented at the behest of the Group Insurance Commission (GIC, the largest health care purchaser in the state), and these programs account for the vast majority of enrollment in tiered products in the state. The six health plans participating with the GIC have all offered products that rely on a claims database aggregated across all of the plans. The aggregated database provides most of the efficiency and quality data used to assess and tier individual physicians or groups. Two additional health plans have smaller products that do not rely on these data.

Several aspects of the tiering methodologies used by the health plans do not conform to the MMS guidelines and raise concerns about the potential for these programs to adversely affect patients and physicians. First, for many specialties, there are few measures and the measure set is not representative of practice. Second, although a few of the plans tier primary care physicians at the group or site of care level, most of the health plans tier individual physicians in both primary care and specialty care. Individual tiering is of concern based on both practical and analytic grounds. Many individual physicians lack adequate sample sizes for accurate assessment, most often of a particular quality measure. In cases where the sample size is inadequate, plans may omit that factor in determining tier placement. This either reduces the weight given to

quality in the calculation or concentrates the quality evaluation on a more narrow set of practice activities. Furthermore, because physicians in a single group can be in different tiers, individual tiering adds complexity for both physician practices and patients. Third, cutoffs are often arbitrary and could lead to misclassification. Finally, the feedback sent to physicians is often inadequate for improvement purposes.

On a positive note, the use of episode-based efficiency measures by the plans is close to being a best practice, and benefits from large sample sizes resulting from the aggregated data, and appropriate data cleaning and handling of outlier cases. Nonetheless, the ETG methodology has yet to be validated for use in physician profiling, in contrast to many of the quality measures.

Summary

Massachusetts health plans have introduced new and revised medical management programs over the past few years. The MMS has created a set of guidelines to serve as optimal standards for the programs. The MMS recognizes the need for improvements in health care delivery and supports effective efforts to reach this goal. Although the health plans meet many of those standards, there are important differences from one plan to another, and some of the guidelines are not met at all. This report details the differences and gaps, and concludes with a series of recommendation to further the benefits that these programs can bring.

Introduction and Project Overview

With the increasing awareness of gaps in health care quality and the continuing rise in health care costs, physicians, payers, and the public are becoming engaged as never before in efforts to improve the delivery of health care in America. Physicians are better organized, often with the help of electronic infrastructure and other systems in their practices. Purchasers of care are more demanding of value in health care, and health plans now use a variety of incentives and tools intended to drive higher quality and more efficient care. The public has unprecedented access to information about health care providers and their demands for additional information are continually growing.

As the health care marketplace evolves, health plans in Massachusetts and elsewhere are instituting new initiatives to address the fundamental issues of costs and quality (i.e., value) in health care. For instance, prior authorization for supplies, services, and medications have been in widespread use for nearly two decades, but new programs such as those directed at high cost radiology are rapidly proliferating and require increasing interaction with physician practices. Further, the majority of US health plans are using some form of pay-for-performance contracting to reward physicians for achieving certain benchmarks or milestones that are said to reflect high quality or efficient care. Finally, tiered network products recently have appeared around the country and in Massachusetts.

As we move forward, controversy remains over the best means to achieving the desired ends. In the absence of empiric evidence, the potential for

ineffective, or even counterproductive, programs remains a serious concern. In addition, collectively these programs have the potential to alter the practice of medicine and the physician-patient relationship by impacting physician autonomy and patient choice, and by imposing additional administrative burdens on physicians and physician practices. The MMS recognizes the need for improvements in health care delivery and supports effective efforts to reach this goal. The MMS also recognizes the potential deleterious effects of these types of programs if not implemented in a manner that is sensitive to the needs of all stakeholders, including patients and physicians.

In an effort to ensure fairness to physicians, to promote the best outcomes for patients, and to encourage reasonable consistency between payers, the Massachusetts Medical Society has taken a proactive approach to payer programs. Over the past few years the MMS has developed guidelines for the appropriate development and implementation of a variety of health plan programs that impact both plan enrollees and participating physicians. We use the MMS guidelines as the benchmark for evaluating the health plan programs.

This report reviews the current landscape of payer programs as they are used in Massachusetts in 2007 and evaluates the extent to which health plan programs conform to the MMS criteria. For each of the major health plans operating in the Commonwealth, this report details key elements of their prior authorization programs, pay-for-performance programs for physicians, and tiered network offerings. Each program is then evaluated against the guidelines promulgated by the MMS. We report both a summary of the current activities in

Massachusetts and identify some of the best practices in use by payers, as well as areas where health plan programs are in conflict with the guidelines.

Methods

We began with the formal published criteria promulgated by the MMS in the areas of prior authorization, pay-for-performance (P4P), and network tiering. The prior authorization guidelines were written by the MMS Task Force on Medical Cost Control. The Task Force did not include any health plan representatives; however, the guidelines have been shared with the payers. The guidelines for tiering and P4P were written by the MMS Committee on the Quality of Medical Practice, which includes twelve MMS member physicians, twelve member physician advisors (i.e., non-voting members of the committee), one resident, and one medical student. Two Massachusetts health plan medical directors were advisory committee members during the development of these principles.

We then divided each of the guidelines into sets of mutually exclusive discrete criteria against which we could judge the performance of health plans operating in the state.

The health plans we evaluated in each area were specified by the MMS and included the three largest health plans in the state (Blue Cross and Blue Shield of Massachusetts (BCBS), Harvard Pilgrim Health Care (HPHC), Tufts Health Plan (THP)), other local health plans with most of their business in Massachusetts (Neighborhood Health Plan (NHP), Fallon Community Health Plan (FCHP), and Health New England (HNE)), and affiliates of national health plans that operate in the state (United HealthCare (UHC), Unicare, Aetna, and

Cigna). The health plans were evaluated only on programs active in Massachusetts.

For each health plan and each set of criteria, we initially gathered publicly available data from the health plan's website and other sources. These data were supplemented with material provided by the MMS including health plan presentations made to the MMS describing details of their programs and other information compiled by the MMS.

We next contacted each health plan and arranged for telephone interviews to confirm information that had already been collected and to gather any missing information. We usually spoke with the Chief Medical Officer or his or her delegate. Associate medical directors and other personnel from network, pharmacy, and quality management were included in many of the telephone calls. After all relevant data were collected; we sent each health plan our documents describing their programs and aspects of their programs relevant to each of the identified criteria. The health plan representatives were asked to review the information and correct any factual errors or misrepresentations. For this report, we relied on the information provided to us by the health plans. We believe that the program descriptions provided are accurate. We were unable, however, to independently verify all details or to test their operational characteristics.

Finally, the performance of each health plan was assessed against each of the detailed criteria for each individual program. For each criterion, we assigned a rating of "met", "partially met", or "not met" based on the objective

assessments of the two independent investigators. In the few cases where there was disagreement, the investigators discussed the assessment until consensus was reached. Neither the MMS nor the health plans contributed to the assessments. We did not formulate an overall assessment or grade for each health plan because there was no clear method for combining the individual assessments into an overall assessment. Similarly, we chose not to rank order health plans. In this report, we summarize our findings and draw attention to areas where the health plans performed well against these criteria, or did not perform well, and we identify individual health plans whose performance against the criteria was better or worse than peers for specific aspects of the criteria.

Detailed factual information and assessments against the MMS guidelines are included as an appendix to this report.

Findings

Prior Authorization Programs for High Cost Radiology

Overview

Utilization management for high cost radiology is widespread in Massachusetts, although not universal. Consequently, differences among health plans relate to details of the implementation of their individual programs. In general, most of the programs are managed by a contracted vendor and, in some cases, the same vendor works with several of the health plans. In these cases, many aspects of program implementation are similar. Most health plan programs apply to CT scanning, MRI, PET, and, for some, nuclear cardiology. Studies ordered from the emergency room and inpatient settings are exempted uniformly from these programs.

One health plan, Fallon Community Health Plan, has a prior authorization program limited to PET scans, which is run internally by the health plan. A voluntary “consultation” program is available through a vendor but is rarely used by physicians. Since PET scans are the least frequently used of high-cost radiology studies, the FCHP program is the least intrusive to physicians. Four of the health plans (Harvard Pilgrim Health Care, Blue Cross Blue Shield of Massachusetts, Neighborhood Health Plan, and United HealthCare) have implemented “prior notification” programs. These programs are similar in appearance to the prior authorization programs implemented by the remaining health plans, but the ultimate determination is non-binding and performance of the study is left to the discretion of the ordering physician. Thus, even in cases

where the study appears inappropriate according to the algorithms employed by the vendor, the study will still be reimbursed by the health plan after all of the appropriate information has been submitted and a registration number has been obtained.

In contrast, Aetna, Cigna, Health New England and Tufts Health Plan require prior authorization for all non-urgent CT, MRI and PET scans and do not reimburse physicians for non-authorized imaging studies.

In several cases, such as with BCBS, HPHC and THP, specific contracting groups are excluded from the program, usually if they have an equivalent system in place that can perform this function internally, or if they are subject to a risk contract. BCBS also exempts some groups that have traditionally performed well. Approximately 50% of BCBS physicians are exempted from the program.

Appropriateness, Communication, and Input

All of the health plans asserted that they had identified both overuse and substantial variation among contracting units prior to implementing their programs. Indeed, these data served as the primary impetus for the programs. This evidence was generally shared with physician contracting units or physician advisory boards prior to the implementation of the programs. No health plan, however, was able to share data on individual physician performance because existing health plan claims data failed to reliably identify the ordering physician.

Each of the health plans sought varying degrees of input from practicing physicians when developing their programs. All of the programs use algorithms

developed based on the literature and expert input. Algorithms are reviewed on a regular basis. In most cases, program development and any program changes were further discussed with clinical advisory boards of the various health plans and stakeholder organizations such as the MMS. HPHC and BCBS in particular sought input from the MMS and made several changes to their program based on this input.

All radiological studies subject to the programs are clearly identified on the health plans' web sites, as are program policies and procedures. Clinical guidelines are available on vendor web sites (e.g., the NIA program used by HNE, HPHC, and THP posts its guidelines at www.radmd.com; the AIM program used by BCBS posts its guidelines on its public website; MedSolutions used by NHP, Aetna and Cigna posts its guidelines in a password protected part of the website that requires registration and therefore is more difficult to access). Although the detailed decision algorithms used by health plan staff are based on the posted guidelines, the health plans and vendors consider these proprietary and do not post them.

Program Implementation

All of the programs place an added administrative burden on physicians and physician practices. Although the programs attempt to minimize the data elements requested, all require a phone call or electronic submission that requires gathering information, which can be time consuming. No additional compensation is offered to account for the added burden. Information can be

submitted by phone or, for some programs, via the internet. Fax and mail are also available for some programs but rarely used. All of the programs conform to standard call center requirements and compliance is actively monitored under all of the contracts. Delays of more than a minute or two are rarely experienced, according to the plans.

All of the health plans exempt inpatient and emergency studies from prior authorization or notification and most of the health plans also have additional provisions for “urgent” studies. These provisions generally involve expedited review, retroactive notification/review, or a combination of both. For instance, if an urgent request requires a higher level “consultation” (i.e., discussion with a reviewing physician), HPHC will arrange for that consultation within 3-5 hours. During business hours, NHP assures a response within the same business day. In addition, for BCBS, HPHC, HNE, THP, and UHC, notification is permitted as late as 2 days after the performance of an urgent study. NHP requires notification within one day. Approvals are granted on the first phone call for the majority of cases and, if not, generally within two business days. UHC’s program stands out as most convenient when a consultation is required. For requests that are not consistent with guidelines and require a peer-to-peer conversation, UHC allows the ordering physician to complete the consultation while on the phone or to select a convenient time during the following three business days to complete the consultation.

For programs that require authorization (as opposed to only notification) the appeals processes for patients conform to federal and state law, as well as

accreditation requirements. Information on appeals is easily accessible and in-person appeals are available upon request although rarely used.

Additional Issues

There are several areas where the performance of the health plans does not generally meet the MMS criteria. First, the MMS criteria espouse the principle that physicians who conform to guidelines should be exempted from the prior authorization program. This has rarely happened. At the start, no individual could be exempted because the health plans could not identify practice patterns without knowing who had ordered an exam. Both HPHC and HNE regularly review performance data and “exempt” high volume physicians who generally conform to guidelines; however, very few physicians are exempted and these physicians still must provide notification about all studies. The other programs do not exempt individual physicians. As noted above, specific contracting organizations are exempted from some of the programs.

Second, individual physicians do not receive actionable reports summarizing their performance in the program. Some health plans (e.g., HPHC, HNE, BCBS, and NHP) share this information with contracting groups, sometimes routinely, and sometimes upon request. Third, denials for THP, HNE, Aetna, and Cigna are generally not issued by actively practicing physicians who regularly treat the condition under question. Rather, denials may be issued by non-practicing physicians from any specialty, but are usually issued by radiologists.

Prior Authorization for Pharmaceuticals

Overview

Pharmaceutical utilization management programs are among the oldest utilization management programs in use by health plans. We examined the application of prior authorization of pharmaceuticals among the Massachusetts health plans. Since each health plan's programs may vary from one drug class to another, we chose a single drug class, oral anti-fungal medications, for evaluation and comparison across the plans. We chose oral anti-fungals because they are expensive and frequently used for treatment of a cosmetic condition (uncomplicated onychomycosis). Oral anti-fungals are, therefore, generally subject to prior authorization.

Of the eight health plans included in this portion of the study, all except HNE and UHC require some form of prior authorization. HNE places anti-fungals on formulary "tier 3" and requires the highest patient copayment, whereas UHC automatically covers the first three months of treatment. After the initial course, UHC allows additional courses of antifungals for certain diagnoses, subject to prior authorization.

Appropriateness, Communication, and Input

Since these drugs are commonly prescribed to treat a cosmetic (and therefore non-covered) condition, evidence of overuse and sharing such evidence with network physicians is less pertinent. Formulary tier placement and prior authorization requirements are usually based on clinical priorities and drug

costs as opposed to variations in use. All of the health plans have a Pharmacy and Therapeutics Committee that includes practicing physicians independent of the health plan to advise the health plan on formulary decisions. Information about pharmacy utilization management programs, including drug classes subject to PA, requisite forms, procedures and coverage policies are available on all of the health plan web sites.

Program Implementation

All of the programs place additional administrative burden on physicians and physician practices, but offer no additional compensation. Although the programs attempt to minimize the data elements requested and the number of agents subject to prior approval, all require the physician or her delegate to submit information. Mostly these programs rely on fax submission, but about half of the health plans allow for data submission by phone and a few allow electronic submission. The latter options are rarely used. Printable fax forms, often drug class specific, are available on health plan web sites. Similarly, clinical criteria for coverage are also available. Fax confirmations are not routinely sent (instead physicians must rely on the electronic confirmation that the fax has been sent). Turnaround time for requests is generally within two days, although several of the health plans maintain a one business day turnaround. Program changes and information are communicated through general and pharmacy-specific newsletters, special mailings, and on the web sites.

For programs that require authorization, patient appeals processes conform to federal and state law as well as accreditation requirements. Information on appeals is easily accessible and in-person appeals are available upon request for most of the health plans (although infrequently used). Individual physician performance in these programs is generally not monitored or fed back and no physicians are exempted from drug prior authorization programs.

Additional Issues

Denials (if part of the program) are generally not issued by actively practicing physicians who regularly treat the condition under question. Rather, denials may be issued by physicians from any specialty, or in some cases, pharmacists. For drug prior authorization programs, data on physician performance are not tracked. Denial lists are maintained or available upon request, but subsequent care for these patients is not monitored for adverse outcomes. As noted above, no physicians are exempted from programs.

Pay for Performance

Overview

All of the local Massachusetts plans have implemented P4P programs aimed at primary care physicians. HPHC, THP and FCHP have PCP-oriented programs that also apply to specialists and are implemented at the level of the contracting unit. BCBS was the only health plan with distinct programs aimed at primary care physicians (The Primary Care Physician Incentive Program (PCPIP)) and specialists (The Group Performance Incentive Program (GPIP)), although the specialist program is also applicable to primary care physicians in multi-specialty groups. Although the three national plans all have programs in other markets, with the exception of a small number of ad hoc arrangements by Cigna, they have not instituted P4P in Massachusetts.

Transparency of the Criteria and Methodology, and Feedback

Most P4P programs used in Massachusetts are aimed at the group or contracting unit and focus on measures applicable to primary care. In many cases, programs are semi-customized such that the contracting unit chooses from a menu of potential measures. For instance, medical groups in the FCHP P4P program select from a list of HEDIS measures on which their performance will be measured. Because these programs are at the group or contracting unit level, local medical directors are responsible for disseminating information and data about the programs to their member physicians. FCHP sends individual

physicians progress reports and a final report that includes a description of the individual measures.

The only plan that disseminates details to all physicians in the network is BCBS's primary care physician incentive program (PCPIP). This is a standardized program that applies uniformly to all PCPs in the BCBS network. A detailed program description including exact measure specifications and thresholds is available on the web and posted prior to the start of the applicable measurement year. This information is updated annually.

With the exception of BCBS, there are no P4P programs aimed specifically at specialists. Most programs, however, include both specialists and PCPs within the contracting units, though most of the quality measures fall in the PCP domain. BCBS's Group Physician Incentive Program is a distinct program that is focused on specialists, and requires that 75% of the payments be directed towards specialists.

The FCHP, HPHC, NHP and THP programs base their quality measures heavily on PCP activities, but the size of the incentives may be greater for specialists (since incentives are allocated often in proportion to total physician payments). Thus, the specialists are at risk for the clinical activities of their PCP colleagues, and the PCPs do not reap the full value earned through the P4P program even though they are most responsible.

It is the view of the MMS that P4P should be used as a tool to both reward physicians and to enhance quality of care. To achieve the latter function requires that programs share goals and data on performance with physicians prior to and

during the measurement year so that physicians have the opportunity to improve upon their performance. With regards to P4P goals and metrics, Massachusetts health plans make this information available to groups and physicians either prior to or early in the measurement year, with performance data for the previous year usually available within the first four months. Not having the data available sooner both diminishes the effect on quality improvement and limits the physicians' ability to achieve the program targets. Further, the data are not always at the individual physician level. As noted above, FCHP sends individualized progress reports and a final report to physicians, and BCBS makes data available to both groups and physicians detailing progress to date. In addition, with the exception of HNE, all of the health plans provide some form of updated performance data to either groups or individual physicians throughout the measurement year (usually on a quarterly basis).

An additional important goal of the MMS is the inclusion of measures that create incentives for physician groups/contracting units to adopt technologies and other processes designed to improve their infrastructure for managing care and improving quality. HPHC and BCBS explicitly include such measures (e.g., EHR adoption, e-prescribing) in their principal P4P programs. Health plans that do not include such measures, including THP and HNE, have additional programs to support infrastructure improvements, whereas FCHP provides specific tools such as registries that can be used to improve care.

Finally, from the physician perspective, it is preferable for programs to rely upon uniform measures with similar specifications, as is done in some parts of

the country such as the California Integrated Healthcare Association (IHA) program. In the case of Massachusetts health plans, P4P programs are not standardized due to health plan competition, contracting unit preferences, and legal reasons. Fortunately, there is significant overlap in what health plans assess in P4P programs. Health plans have an opportunity to consolidate the measures that they use, preferably with the cooperation of the MMS and medical specialty societies.

Monitoring

As with any performance measurement and reward program, there is the potential for P4P programs to result in unintended consequences.

Massachusetts health plans are not aware of any unforeseen consequences, but most of the health plans do not have specific monitoring programs in place other than voluntary feedback from physicians and members. Two of the health plans (BCBS and HPHC) specifically monitor quality performance across practices and geographic areas to ensure that there are no untoward consequences for specific populations or geographic areas.

Magnitude and Sources of Funding

In general, P4P incentives appear to be of sufficient size (i.e., 5-10% of total compensation) to engage physicians in performance improvement, although the incentive size varies from program to program.

The sources of funding for P4P (whether taken from the existing reimbursement pool or designated as new additional funding) is to some degree an accounting issue. Although the MMS guidelines specify that funding for P4P programs should not be a simple redistribution of current physician pay, the distinction is often not clear. For the majority of the health plans, P4P payments are negotiated as part of the overall contract, so whether this represents additional resources or a redistribution from what would have been paid is difficult to discern. Three health plans (BCBS, NHP, and HNE) use additional funds beyond those contractually specified, although the availability of these funds ultimately is dependent upon overall health plan performance.

Tiered Networks

Overview

Eight Massachusetts health plans offer tiered network products based on measurements of quality and efficiency*. Most were created in response to the Group Insurance Commission, the largest health care purchaser in the state, which required all of its participating plans to develop a tiered product. Some GIC participating plans also offer their tiered products to non-GIC customers. BCBS and Cigna, which do not participate with the GIC, offer tiered products as well. Each of the health plans creates their tiered network on the basis of measures of the quality and efficiency of their network physicians. Based on the measurements, physicians (or physician groups) are placed into one of two or three tiers, each of which generally has a different copayment level. The six GIC plans have many common features in their tiered products, since they participated in a combined quality and efficiency dataset based on the pooled administrative claims of all six health plans. HPHC, THP and HNE add some health plan-specific data when creating their tiering.

Strengthen the Patient-Physician Relationship.

All the tiered products may adversely affect doctor-patient relationships, for example by disrupting a pre-existing clinical relationship by charging the patient a higher copayment to remain with their doctor. In addition, with

*Most Massachusetts health plans also offer simple tiered products that are based on provider characteristics such as primary care vs. specialist physician. These plans are not included in this discussion.

individual tiering, offices may need to collect different copayment amounts for different doctors in the practice. This is burdensome to both the practice and the patient, and it is most problematic for practices such as primary care and cardiology where patients are often shared or seen by multiple physicians within the group. Tiering at the group level makes it easier for patients to see a practice partner of their regular physician without adding any administrative burden or confusion. FCHP encourages choosing a preferred (“Value Plus”) PCP by applying a lower copayment not just to PCP visits, but also to specialist referrals from the PCP. This may encourage coordination of care among the patient’s doctors. Further, to reduce the administrative burden to physicians, FCHP has all offices collect the same copayment. The plan then refunds the difference to the patients later.

For primary care, HNE, FCHP, NHP, Cigna, and BCBS tier at the group or site level (in some cases, a “site” may include just a single physician). UniCare tiers primary care physicians at the individual level. HPHC and THP do not tier PCPs.

For specialists, HPHC and THP tier some specialists on an individual basis and others on a group basis. BCBS does not tier specialists. HNE, NHP, Unicare and Cigna tier specialists at the individual level. FCHP tiers specialists at the group level only.

Most plans revise the tier placements annually. UniCare allows a physician who has earned preferred tier placement to remain preferred for two

years, even if they fail to meet the criteria in the second year. This may be less disruptive to patient and physician alike.

When used appropriately, the quality measurement and public reporting activities associated with tiering efforts may promote better clinical relationships as well. Since the quality measures (such as HEDIS) reward behaviors of both physicians and patients, an implicit incentive is created for improved physician-patient collaboration. THP goes further and includes patient satisfaction scores as part of its measurement system. Greater satisfaction scores are likely correlated with better clinical relationships. A potential, yet unconfirmed, adverse effect on clinical relationships is the perceived pressure on physicians to discharge patients who are non-adherent (or incur unusually high costs of care) from their panels, in order to avoid lowering the physician's quality and cost ratings.

Involve Physicians in the Design and Implementation of All Programs

The six GIC-participating plans all agreed to use pooled ETG efficiency data early in their design process. Thus, none of the six allowed any substantial input from physicians into how the efficiency ratings would be performed. Similarly, none of the plans were very responsive to physician concerns on certain "business decisions" in the design of the programs (e.g., copayment levels or the proportion of physicians earning preferred tier placement).

With the exception of BCBS and Cigna, all the plans made outreach efforts to involve network physicians and the MMS when creating their tiered

products. HPHC sought input from multiple stakeholders and incorporated the feedback received into program modifications. THP also sought extensive input, and allowed an independent expert advisory committee to determine many key aspects of its program. Most of the aforementioned input, however, related to quality measurement because none of the GIC plans had much flexibility with respect to the ETG methodology. HNE incorporated stakeholder input and has already begun the process of soliciting input for changes in next year's product. BCBS created its program without outside physician advice; however, they were responsive to MMS suggestions for modifications. FCHP, NHP, and UniCare all discussed their programs with standing physician advisory committees, but did not seek additional input from the MMS or others.

The criteria used for measuring quality and efficiency were circulated in advance by HPHC, THP, HNE and UniCare. While the other plans also shared the measurement criteria, they did not do so prior to releasing their physician ratings.

Share Data with Physicians/Practices Prior to Public Release

The short timeframe for implementation faced by the six GIC-participating plans limited their efforts to give their network physicians detailed data well in advance. In contrast, BCBS provided performance data to groups a minimum of three months prior to the launch of their product.

Of the GIC plans, THP was most successful in sharing quality data in advance with physicians and allowing a (brief) opportunity for data checking,

correction, and appeals prior to finalizing or releasing their tier placements. THP provided the component scores of quality and efficiency and backing data (limited to THP-specific data, such as member satisfaction and ETG data on THP members).

FCHP shared detailed data with many of their network physicians though not in advance of public announcement of tier placement. FCHP met with most physicians after announcing tier placement but early enough to allow physicians an opportunity to review and correct data or calculation errors before the program launched.

HNE shared detailed ETG data with physicians at the practice site level. HNE allowed physicians to appeal tier placement around the time the proposed tier placement was made public.

A common problem for the GIC plans is that the use of pooled (and therefore de-identified) data made patient-level data unavailable for sharing with physicians (except for data on the patients enrolled in that particular plan).

While all of the plans share at least summary information explaining a provider's score, none suggested specific behaviors for physicians to improve their scores. Of the plans, THP made the most complete information available. THP provided ETG reports showing individual physician performance in six different categories of care (e.g., pharmaceutical use) for each physician's top seven ETGs (compiled from THP members only, not the pooled Mercer/GIC dataset). THP also shares HEDIS data with physicians and groups. However,

THP's shared data does not cover the full spectrum of measures included in their tiering calculation.

The other health plans release available information upon request, but not routinely. All GIC participating plans do not share the common de-identified quality and efficiency data. BCBS and Cigna also do not routinely share data, although for BCBS, the quality measures overlap with those used in other quality measurement and reward activities for which physicians receive detailed performance reports. BCBS also makes their performance data available to physicians on the web.

HPHC and FCHP share detailed data and patient lists for HEDIS measures, but not other measures used in their tiering program. A great performance improvement opportunity exists if the pooled GIC data, with patients individually identified, could be efficiently disseminated to Massachusetts physicians and groups.

Support Clinically Important and Sound Performance Measures

The quality and efficiency measures were selected by the plans based mainly on availability, feasibility, and validity grounds. Potential unintended consequences may result, including physician misclassification, and neglect of unmeasured activities.

Quality Measurement

The GIC health plans tier physicians or groups in as many as 15 different specialties. Unfortunately, the development of quality measures has focused on a handful of specialties, leaving some specialties with few validated measures. For example, the quality measures used by most health plans for tiering primary care (internal medicine, family practice and pediatrics) are reasonably valid and representative of a variety of the clinical issues within those specialties. Even within primary care, however, available measures are not representative of all of the activities performed by primary care physicians, and under-represent diagnostic and other skills that are difficult to measure. In addition, the reliance on measures derived from medical claims precludes the inclusion of outcome measures aimed at the successful management of chronic medical conditions (e.g., control of diabetes or high blood pressure). All plans include accepted measures of preventive care in their primary care measure sets. This may encourage broader use of these sometimes underused services.

In contrast, the measure sets available for most sub-specialties often rely on a small set of measures that are not representative of the most important activities of that specialty. Dermatology is a good example, as the Resolution Health, Inc. (RHI) measure set used in the GIC process includes only a single measure related to drug safety for the teratogenic acne medication isotretinoin (Accutane[®]). This measure should rarely, if ever, not be met. Relying on such a narrow measure is unlikely to accurately identify quality practice.

UniCare omits quality measures for specialties where there are fewer than 5 RHI measures available. This avoids some problematic quality measures, but it leaves the tiering for those specialties to efficiency measures alone.

For certain specialties, some plans have adopted different strategies for tiering that rely on supplemental data from the health plan. For instance, HPHC enriched their ratings by gathering quality measures from several sources. In addition to the quality measures based on Mercer/GIC pooled data using RHI measures, HPHC uses ProfSoft's SAGE measures, and measures from AHRQ and NQF. THP, after an expert panel found many of the specialty-specific measure sets lacking, adds member satisfaction data (for some specialties) plus information on the use of EMRs, e-prescribing, and certification with NCQA's Physician Practice Connection. The additional measures help to broaden the assessment and may add to its accuracy.

All plans use quality measures that are evidence-based, but some are not broadly accepted or endorsed by national organizations such as the NQF. Further, some measures have not been validated for use with claims data. The measures used may lead to excessive misclassification of physicians. Among specialties, this is particularly problematic for general surgery, infectious diseases, neurology and ophthalmology, and it is least problematic for cardiology.

The specialist measures could be improved by soliciting new measures from medical specialty societies and other stakeholders. These could supplement or replace measures currently in use.

Efficiency Measurement

Efficiency measurement is not nationally standardized. The efficiency measure used by all six GIC plans is Symmetry's Episode Treatment Groups (ETGs). ETGs have become a de facto standard, despite not being validated for use with individual physicians. ETGs rely on accurate diagnostic coding by physicians and may mislead if coding is not consistent across the physician network. They are widely used by health plans nationally and have been used in MA for several years. The GIC plans follow commonly used procedures and apply appropriate data checks.

BCBS measures risk adjusted total costs using Diagnostic Cost Groups (DxCG), which is widely accepted and also is a validated method used by CMS in the Medicare program, but not a national standard.

Insure Adequate Sample Sizes to Support Meaningful Analysis of Data

The plans have generally reasonable statistical techniques to avoid errors from inadequate sample size, although there will always be important tradeoffs between the sample size requirements at the individual physician level and the proportion of a plan's network that qualifies for measurement. Measurement at the group level offers the advantage of greater statistical power in any comparison. Even where individual tiering is used, a set a minimum number of observations, usually thirty, is required to create a rating. Although some argue that this threshold should be higher, there is no agreed upon standard. The use

of pooled data from six plans gives great statistical power to both quality and efficiency measurement. The dataset used for efficiency measurement includes three years' of data, further enhancing the statistical power. BCBS has a large membership such that they have reasonably sufficient power to measure individual physicians. None of the health plans incorporate advanced statistical methods to control for clustering (the degree to which patients measured within the same practice are similar) within groups and physicians. These methods, while appropriate, are not widely used outside of academia.

Support Meaningful/Accurate Data Analysis

One tradeoff from maximizing statistical power with three years' data is a loss in the timeliness of the data. For example, among the six GIC plans, if a physician this year markedly improves her performance, the improvement will be diluted with two other years' data. In addition, the pooling process takes time, so that the claims data behind the 2007 tiering dates from 2003 to 2005. This is far from ideal but also not likely to change in the near future. The trade-off for data completeness, aggregation, and checking is a delay in the timing of any analysis. This unsatisfying tradeoff highlights one reason behind the Medical Society's policy opposing tiering at the individual level, although group level tiering would not necessarily use more recent data.

None of the plans makes adjustments for patients' health status, adherence, or socio-economic factors. This could promote disparate treatment. However, since the ratings are derived only from the commercially-insured

population (generally non-elderly with an employed household member), the effect of such factors is likely to be small.

The health plans vary greatly in their processes for data review, feedback and correction by physicians. None of the plans allows more than about one month for physician feedback, and some of the health plans, such as Unicare, only respond to physician-initiated requests for review. For some plans, the opportunity for feedback comes after tier placements are already published. HPHC alone does not allow any changes to its ratings once they are released.

The performance rating of a physician is highly dependent upon the measures used. This is certainly true for quality measures based upon services rendered to individual patients of a physician. However, some of the sub-specialty quality measures may be more reflective of care rendered by PCPs than the specialists they purport to evaluate. For example, several plans measure otolaryngologists on the care rendered to their patients with pharyngitis, even though much or most of that care may be rendered by the patients' primary care physicians. An otolaryngologist's measured performance may have more to do with the PCPs who refer to him than to the care he renders to his patients.

The ETG (and DxCG) efficiency measures explicitly include care rendered by all physicians involved in the case. But the ETG is attributed only to the physician who rendered most of the professional services and who was presumably involved to some degree in all aspects of that patient's episode of care. Because multiple physicians' services are included in an episode attributed to the rated physician, ETGs violate the MMS guideline that the measurement

should only capture behavior of the physician being rated. Despite this, ETGs are widely accepted, and they are applied with a reasonable methodology by the plans.

Most of the plans use “dummy” pricing for services, thus insulating physicians from the effects of their unique contractual reimbursement rates. The advantage of dummy pricing is that the efficiency measure will reflect only differences in utilization of services or supplies, not pricing. It allows physicians to focus on their practice patterns, not their contracts. THP, HPHC, and Cigna do not use dummy pricing, instead re-pricing each episode to reflect its actual cost to the payer. This is intended to keep contractual pricing in the consciousness of physicians, and to provide an incentive not to seek higher fee-for-service reimbursement. None of the plans adjust for local wage rates, but these do not vary greatly within most of MA.

Each tiered product, in the end, makes a firm determination of whether a physician is preferred or not. While each plan uses somewhat different methods and thresholds, any cut point chosen is at least partially arbitrary. The more arbitrary or random the threshold is, the greater the risk of misclassification error. FCHP, THP and HNE sought a “bright line” or “natural cut point” in the data wherever possible, rather than keeping to a strict numeric threshold. This is a simple technique that reduces arbitrariness to some degree, although there is no way to avoid misclassification when choosing any cut points. Each plan also made a business decision regarding what proportion of physicians to allow into

the preferred tier. This varies between plans, and whatever value is chosen is arbitrary.

Support Transparency of All Quality Measures and Approaches to Data Analysis

HPHC and THP represent best practices for transparency regarding the methodology they used. Each published detailed methodology documents on its website. Both make similar information available to members via the web as well. UniCare posts detailed information for physicians but has more limited information available to members. FCHP shares detailed information in person with most physician groups. However, some FCHP physicians receive no information, and little information is available to physicians or members via their website. The other health plans provide more limited information.

Explicit Consideration Should be Given to Unique Circumstances of Practices that May Require Special Attention in Quality Monitoring

All tiered products make analytic accommodation for new and small practices, which will have less data available for measurement purposes. Most of the health plans automatically default these practices to the non-preferred tier; however, FCHP, NHP, and BCBS default these practices to the preferred tier. None of the health plans make special accommodations for practices that have very different practice demographics, such as a community health center or rural practice.

Support Uniform Reporting Formats

Each of the plans expressed support of the notion of uniformity, and six collaborated in the joint dataset. However, each plan currently has its own format for sharing data with physicians and for communicating about its tiered product. Each plan also has its own format for sharing tiering information with the public. An opportunity exists to collaborate to create a common format for feedback to physicians and to the public. The joint GIC process would be a natural starting point. A newly created physician advisory committee with the GIC, health plans and MMS representatives concluded discussions with the plans that resulted in agreeing to such a common format for use next year.

Minimize Unintended Harmful Consequences of Quality Monitoring and Public Reporting of Data

While each of the plans investigates specific complaints received from members or physicians, none systematically monitors for potential adverse consequences.

Summary

Eight health plans offer tiered products in MA. Most have successfully obtained some degree of network physician input into their programs, used reasonable analytics given the requirements of their major client, and were reasonably transparent about their process. Most, but not all, also used techniques to maximize statistical power and allowed some kind of feedback and

correction process. The substantial efforts made by some plans should be recognized and encouraged. Unfortunately, all suffer from trading off the use of more timely data in exchange for using complete, multi-year data and taking the time to clean it prior to analysis. All (except BCBS) tier at least some physicians at the individual level, which raises methodological concerns, and often causes administrative hassles in group practice offices.

The breadth and validity of the quality measures used is still limited for many specialties. Efforts to adopt new validated, accepted measures should be pursued. A few plans have QI tools available in the form of near-real-time feedback to physicians. These efforts could be emulated and further expanded. Adoption of common measures, techniques and formats may be beneficial to physicians, health plans and patients alike.

Conclusions and Recommendations

Massachusetts health plans have expanded their use of quality and utilization management tools over recent years. In some cases (like prior authorization for pharmaceuticals and radiological studies), they come close to meeting all of the MMS criteria, although most physicians would much prefer to not have these programs altogether. Other programs, such as tiered network products, fall short of the ideal. We suggest some possible improvements for each of the programs.

In the final analysis, physicians may continue to feel burdened by health plan programs with legitimate quality improvement and cost reduction goals. The cumulative effect of the many different programs, even when well-run, is to place significant administrative burdens and other pressures on practicing physicians. We believe that collaborative dialogue between physicians and payers can bring continuous improvement to the quality and efficiency of these programs.

Specific Recommendations:

- Every effort should be made to exclude highly performing physicians from prior authorization programs that require time and effort on the part of physicians that could be devoted to patient care. Similarly, groups that can demonstrate similar capabilities should be excluded from programs.
- Health plans should regularly review their programs that add administrative burdens to physician practices. Programs should be

considered for modification or elimination if they are not clearly meeting their cost or quality goals.

- Although empirical data are lacking, prior notification appears to be almost or as effective as prior authorization. Such programs are more sensitive to physician needs, while achieving similar cost savings. Health plans should consider changing prior authorization programs to prior notification.
- Each health plan should publish detailed specifications of their programs that are available to physicians and patients alike on their web sites.
- P4P programs should link the incentives more directly to the physicians whose behavior is being measured.
- Appropriate statistical methods, such as adjusting for clustering within individual physician's patients, should be used by all health plans when measuring physician performance.
- To facilitate quality improvement efforts of physicians, health plans should expand the use of individualized real-time feedback to physicians. In addition, efforts to enhance the usability of data including the adoption of common measures, measurement techniques, and common feedback formats should be pursued where feasible.
- The lack of patient-level data limits the usability of the aggregated efficiency data being used in the GIC tiering process. A great performance improvement opportunity exists if the pooled GIC data, with patients individually identified, could be efficiently disseminated to Massachusetts physicians and groups, as is currently contemplated for

the next cycle of the GIC tiering process. Similarly, the proposed Massachusetts all-payer dataset could be used for this purpose.

- To improve the breadth and depth of quality measures in specific specialties, the health plans, potentially through the MMS, should solicit input from specialty societies and expand efforts to identify and adopt validated accepted measures. For specialties with inadequate measures, the benefit of the program with existing measures should be weighed against the potential for deleterious effects.
- All health plans should adopt formal appeals processes for tier placement. Physicians should receive their data and have sufficient time to submit corrections and appeals before the proposed tiering is made public or tier placement is finalized. Specific plans discussed in this report could be used as models. If necessary, it is preferable to use less timely data but to allow more careful review, correction and appeal processes.
- With current methods, there are substantial limitations to our ability to measure the performance of individual physicians. Consequently, for tiering programs and most P4P programs, consideration should be given to restricting measurement to physician practice sites or groups. Implementing this recommendation does not require health plans to tier at larger contracting units such as Independent Practice Associations.
- Systematic efforts that span different quality and utilization management programs should be put in place for each plan to monitor the potential

unintended consequences of these programs on disadvantaged populations and the physicians who provide care for them.

Acknowledgements

The authors thank the staff of all of the health plans included in this report for their gracious and extensive cooperation with this project.

We also greatly appreciate the work of Rachel Henke, Ph.D., for her invaluable assistance.