

Contract No.: RFP03-06-MedPAC/E4016631
MPR Reference No.: 6355-300

MATHEMATICA
Policy Research, Inc.

**Health Plans' Use of
Physician Resource Use
and Quality Measures**

Final Report

October 24, 2007

*Timothy Lake
Margaret Colby
Stephanie Peterson*

Submitted to:

Medicare Payment Advisory Commission
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Project Officer:
Jennifer Podulka

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Timothy Lake

ACKNOWLEDGMENTS

We would like to thank MedPAC staff—Niall Brennan, Jennifer Podulka, and Megan Moore—for their active participation in site visits, their guidance in shaping the design of the study discussed in this report, and their comments on an earlier draft of the report. We would also like to express our appreciation for the health plan staff and physician practice representatives who met with us during the visits and provided the information and insights reflected in the report. Finally, we would like to acknowledge Mary Laschober, Barb Geehan, and Felita Buckner at MPR for their assistance with the review, editing, and production of this report.

CONTENTS

Section	Page
	EXECUTIVE SUMMARY v
A	BACKGROUND AND METHODS..... 1
B	KEY FINDINGS 3
	1. Market Context for Resource Use and Quality Measurement 3
	2. Health Plan Measurement Approaches 9
	3. Physician Reaction 14
C	LESSONS FOR FUTURE MEASUREMENT EFFORTS..... 19
	APPENDIX A: INTERVIEW PROTOCOLS
	APPENDIX B: SEATTLE SITE SUMMARY
	APPENDIX C: BOSTON SITE SUMMARY
	APPENDIX D: AUSTIN SITE SUMMARY
	APPENDIX E: CLEVELAND SITE SUMMARY

EXECUTIVE SUMMARY

In recent years, measures of health services resource use have been developed for physicians and other providers to assess efficiency of care, primarily relying on tools commonly known as “episode groupers.” Resource use measures are often combined with measures of quality of care. Applying episode groupers for physician resource use measurement involves: (1) identifying episodes of care comprised of clinically-related healthcare claims data (including hospital, physician, pharmacy, laboratory and other types services) over a defined period of time, (2) attributing episodes to a physician or group of physicians, and (3) comparing the actual costs of episodes to their expected costs for each individual or physician group.

To investigate the uses of these episode grouper-based measures in the private sector, staff from Mathematica Policy Research, Inc. and the Medicare Payment Advisory Commission conducted site visits to health plans operating in four markets around the country: Seattle, Washington, Boston, Massachusetts, Austin, Texas and Cleveland, Ohio. The health plans interviewed represent a mix of national managed care companies, Blue Cross/Blue Shield plans, and local health plans. During our site visits, we also met with representatives of local medical societies, physician and/or practice managers at 3 to 4 physician practices per site, and staff from health care purchasing groups. Because we selected health plans and physician practices non-randomly, results from this study are not necessarily representative of experiences nationwide, but instead provide the perspectives of a diverse group of early adopters and users of these measures in selected markets.

FINDINGS IN BRIEF

Uses of Performance Measures. The health plans we interviewed use resource use and quality measures together in profiling the performance of physicians. In development of their measurement approaches, health plan staff devote considerable time addressing technical measurement issues including: choosing and refining episode groupers, defining adequate patient sample sizes per physician or practice, developing methods for attributing episodes to particular physicians, and establishing appropriate benchmarks for comparisons of performance.

Health plans are in early stages of investing in methods for communicating results of measurement efforts to physicians. Communication about how physicians fared on particular measures is now usually limited to a brief letter indicating their overall scores and explaining how their performance affects their status within a plan’s network. Health plan staff agreed that investment in more in-depth methods for explaining results to physicians will be important for future efforts to constrain use of resources and improve quality of care.

Among the health plans we visited, performance measures are used in four main activities in the marketplace: (1) assignment of physicians to network “tiers” based on their performance with differential copayments charged to consumers based on choice of physician within a tier, (2) development of lower premium insurance products with smaller “high-performance” networks, (3) providing feedback reports to physicians, and 4) reporting to consumers about physician performance, often referred to as “transparency” efforts.

Physician Reaction. Physician reaction to use of these measures has been mixed, and appears to fall into one of three categories. First, many physicians we interviewed had fairly limited familiarity of specific health plan efforts. Despite limited familiarity, many expressed skepticism about the benefits of or need for specific efforts to measure costs of care.

A second category consisted of providers who have somewhat greater awareness of specific health plan efforts to measure resource use but who are dissatisfied with specific aspects of these measurement efforts. Physicians in this category raised concerns about the level of communication they had received from health plans. For example, some physicians said they received a letter reporting their performance score, but received little additional information explaining how the scores were developed or why their score deviated from the benchmark.

A third category, including a large multi-specialty group and a smaller primary care group practice, also were fairly familiar with measurement efforts like the second group, but expressed more positive views about measurement effort and are more active users of the information, considering performance measurement a key component in their market strategy.

LESSONS FOR FUTURE MEASUREMENT EFFORTS

The private health plans we visited have multiple years of technical experience implementing physician resource use and quality measures, particularly with the use of episode grouper tools to compare physician costs of care. However, most are still in the early stages of determining the best ways to use these measures in their local markets. The Medicare program can draw on their experience for considering potential future uses of these measures in several ways.

Implementing Resource Use and Quality Measures Together. For each of the health plans we visited, physician resource use measurement has been implemented in combination with process-of-care quality measurement, usually through an approach in which physicians who first meet quality standards are then rated on cost. Health plans believe that resource use measures that are not accompanied with quality measures may not be well received by physicians and may remind physicians of prior “economic profiling” efforts that attempted to remove physicians from plan networks based solely on costs.

Exploring Multiple Uses of Performance Measures. Health plans visited during this study are in the early stages of exploring different uses of performance measures. Efforts to date focus on simple feedback reporting to physicians and on encouraging consumers to use high performing network tiers through public reporting and limited financial incentives for consumers. Health plan staff note that these efforts are likely to evolve over time as they assess the effects of different approaches, including enhancements to physician feedback, public reporting, and/or pay-for-performance.

Providing Actionable Feedback to Physicians. Stakeholders in our interviews agreed that providing actionable and well-accepted feedback to physicians on their patients’ service use compared to those of their peers, along with quality measures, is key to constraining resource use in the future. Health plans have begun investing resources in physician communication efforts, but the degree of communication varies among plans, and many physicians remain unfamiliar

with each plan's measurement approach. Physicians also raised concerns about the lack of standardization in approaches and the inability to “drill down” to find actionable information.

A. BACKGROUND AND METHODS

In recent years, measures of health services resource use have been developed for physicians and other providers to assess efficiency of care, primarily relying on tools commonly known as “episode groupers.” Resource use measures are often combined with measures of quality of care. Applying episode groupers for physician resource use measurement involves: (1) defining episodes of care comprised of clinically related healthcare claims data (including hospital, physician, pharmacy, laboratory and other types services) over a defined period of time, (2) attributing episodes to a physician or group of physicians, and (3) comparing the actual costs of episodes to their expected costs for each individual or physician group.

Over the past couple of years, the Medicare Payment Advisory Commission (MedPAC) has been testing episode grouper software packages on Medicare fee-for-service claims (MedPAC 2006). This report complements the Commission’s quantitative analysis of such grouper tools with qualitative insights from a study of private health plans’ use of resource use measurement. MedPAC’s ultimate goal is to assess how these tools might be used effectively within the Medicare program.

For the study, staff from Mathematica Policy Research, Inc., and MedPAC conducted site visits to health plans operating in four markets around the country: Seattle, Washington (June 11-12, 2007); Boston, Massachusetts (June 18-19, 2007); Austin, Texas (June 26-27, 2007); and Cleveland, Ohio (July 18-19, 2007). The health plans interviewed represent a mix of national managed care companies, Blue Cross/Blue Shield plans, and local health plans. First, we developed a list of targeted health plans that represented a mix of geography, type, and ownership, with some preliminary information about their use of resource use and quality measures based on document or web site reviews. Through pre-screening telephone calls with health plan staff, we selected plans with a diversity of approaches for measuring resource use and

uses of those measures in the market place. We conducted semi-structured interviews with multiple staff at each selected plan that typically included a medical director, quality improvement staff, and/or information technology staff. During our site visits, we also met with representatives of local medical societies, physician and/or practice managers at several physician practices, and staff from health care purchasing groups. Site visit interviews were supplemented with telephone interviews with additional health plans and physician practice representatives. In total, we interviewed representatives from 15 physician groups, 5 medical societies, 2 purchasing coalitions, and 4 health plans.¹

Key research questions addressed during the site visits included (see Appendix A for interview protocols):

- What approaches are health plans using for measuring resource use?
 - What factors lead them to choose these approaches?
 - What technical issues have they faced in applying the measures, and how have these issues been addressed?
 - How are quality measures used in combination with resource use measures?
- How are health plans using measures in the market place, including for physician payment, network development, and assisting consumer decision-making? What effect have these approaches had on the local market?
- How have physicians reacted to health plans' use of resource measures?
 - How aware are they of the measures, and do they view them as accurate and valid?
 - How do they use measures in their efficiency improvement or cost containment efforts?

Because health plans and physician practices were selected non-randomly in four markets, information gathered in our interviews does not necessarily reflect the experiences of all health

¹ We also conducted five additional screening interviews with staff at other health plans who were implementing resource use and quality measures. The information from these interviews was generally consistent with our findings from the four visited plans, but the results discussed here are limited to the four visited plans.

plans or physicians nationwide. Instead, they provide impressions of the activities and experiences of a select but diverse group of stakeholders who were relatively early adopters of resource use and quality measures. In the remainder of the report, we present our findings related to each of these questions, and conclude with discussion of implications for Medicare's potential use of efficiency and quality measurement efforts.

B. KEY FINDINGS

Cross-cutting key findings are described below. Our main findings from each site are summarized in Table 1; more detailed results are provided in Appendices B, C, D, and E.

1. Market Context for Resource Use and Quality Measurement

Health plan efforts to measure physician performance on resource use or quality are largely driven by purchaser demands for greater health care value. At two of the site visits (Seattle and Cleveland), we visited large managed care companies that were implementing approaches in multiple markets across the country in response to demand from large national purchasers. In the Austin market, a dominant BCBS plan was implementing its approach statewide in response to local large purchaser demand. In the Boston market, measurement efforts were driven by one large purchaser.

Current measurement approaches are relatively new (two to five years old), but many plans have built on earlier approaches. In each of the markets we visited, multiple health plans (not just the visited plan) were implementing quality and resource use measurement processes.

The visited markets were generally in the early stages of using resource measures to assess provider efficiency, with fairly limited applications to benefit design, payment, or provider selection. Health plans in one market, though, were in the process of adjusting their approaches in response to negative market reactions to earlier, relatively aggressive use of the measures for purposes of provider network development and benefit design. Limited health plan use of

Table 1. Site Visit Summary Table

Aspects of Measuring Resource Use and Quality	BCBS	HPHC (GIC) ^a	Aetna	UnitedHealthcare
Overview				
Year began using physician resource use measurement tools	2002	2004	2002	2005
Episode grouper or other measurement product(s) used	Symmetry ETG Medstat MEG	Symmetry ETG	Symmetry ETG Cave Grouper	Symmetry ETG Symmetry EBM Connect Anchor Target Procedure Grouper (ATPG)
Quality measures used in conjunction with resource use measures	Measures from Health Benchmarks, Inc. (HBI), covering 25 specialties with a total of 36 quality measures.	Measures from Resolution Health Inc. (GIC's quality analysis contractor), HEDIS, AHRQ, and SAGE.	Eight quality measures per specialty that have been endorsed by their profession.	Measures determined by national disease and specialty standards or by plan's Scientific Advisory Boards.
Technical Specifications				
"Gated" approach vs. combination approach ^b	Combination approach: 1) efficiency measures (Risk Adjusted Cost Index [RACI] score) and 2) quality measures	Gated approach: 1) quality, then 2) efficiency	Gated approach: 1) sufficient volume of episodes, then 2) quality, then 3) efficiency	Gated approach: 1) quality, then 2) efficiency
Individual provider and/or group level measurement	Group (Tax-ID) level measurement.	Individual provider-level measurement, rolled up to group-level for tier designation for some specialties	Individual provider-level measurement, rolled up to group-level for tier designation	Both. Individual provider-level measurement, however aggregate group-level score may override individual-level measurement if higher (as long as the provider did not fail).

Table 1 (*continued*)

Aspects of Measuring Resource Use and Quality	BCBS	HPHC (GIC) ^a	Aetna	UnitedHealthcare
Specialties profiled	Most specialties, with some exceptions (such as neonatologists)	9 specialties 5 tiered at individual level: Allergy, general surgery, neurology, ophthalmology, otolaryngology 4 tiered at group level: Cardiology, dermatology, gastroenterology, orthopedic surgery	12 specialties Cardiology, cardiothoracic surgery, gastroenterology, general surgery, neurology, neurosurgery, obstetrics and gynecology, orthopedics, otolaryngology/ETN, plastic surgery, urology, vascular surgery	16 specialties and primary care physicians. Allergy, cardiology, cardiothoracic surgery, endocrinology, family medicine, infectious disease, internal medicine, pediatrics, nephrology, neurology, oncology, orthopedic surgery, pulmonology, rheumatology, obstetrics and gynecology, neurosurgery
Attribution method	Episode assigned to physician who bills the greatest total Relative Value Units (RVUs) for a given episode, as long as the physician has a minimum number of RVUs. When no physician is identified by RVUs, episode is attributed to the physician billing the greatest number of outpatient evaluation or management services for the episode, as long as the physician has a minimum number of outpatient E&M services. When no physician is identified by either of the above, episode is attributed to the physician with the highest allowable cost included in the episode.	Episode assigned to physician with the highest amount of claims dollars, as long as physician is responsible for at least 25% of the episode fees charged. If no physician has at least 25% of the claims dollars for the episode, the episode remains unassigned.	Symmetry ETG: episode assigned to physician with majority of claims dollars included in the episode, or to surgeon if a surgery occurs. Cave Group: Episode assigned to each physician with more than 20% of claims dollars included in the episode.	For non-proceduralists: episode assigned to physician with majority of claims dollars included in the episode. For proceduralists: episode assigned to physician who submitted the claim for the interventional procedure.
Minimum number of episodes required for profiling a physician	30 episodes	30 episodes	20 episodes	20 episodes (for procedural) 10 episodes (for non-procedural)

Table 1 (*continued*)

Aspects of Measuring Resource Use and Quality	BCBS	HPHC (GIC) ^a	Aetna	UnitedHealthcare
Percentage of potentially eligible physicians profiled	72-74%	65%	70-75%	69%
Addressing outliers	Exclude some low-volume episodes for some specialties	Truncate outlier episodes at the 2nd and 98th percentiles for cost for each episode. Exclude some low-volume episodes for some specialties.	Exclude most volatile episodes.	Exclude outlier episodes that fall below the 5 th percentile for cost for each episode. Truncate outlier episodes that exceed the 95 th percentile for cost for each episode.
Methods for benchmarking/ranking	Physicians with scores above certain efficiency and quality thresholds are selected.	Quality—top 90% Efficiency—top 25-30%	Quality—top 95% Efficiency—top 50% as designated by both Symmetry and Cave ETGs.	Quality—non-proceduralists must achieve 70% score on Symmetry EBM Connect. Proceduralists must meet or exceed average on metrics designated by Scientific Advisory Board. Efficiency—top 50%
Price standardization methods	No price standardization	GIC uses price-neutral analysis; HPHC applies contracted rates to GIC data	No price standardization	No price standardization
Data source and/or aggregation (including number of years of data)	Two years of claims data	Three years of claims data. Claims data aggregated from all six GIC health plans.	Two years of claims data	Two years of claims data

Table 1 (*continued*)

Aspects of Measuring Resource Use and Quality	UnitedHealthcare			
	BCBS	HPHC (GIC) ^a	Aetna	UnitedHealthcare
Key modifications over time	<p>Changed from Symmetry ETG to Medstat MEG in 2004.</p> <p>Added quality measures in 2007.</p>	<p>Modification of outlier methodology.</p> <p>Exclusion of some episodes from some specialties.</p> <p>Weighting of most recent year of claims data.</p>	<p>Addition of Cave grouper in 2005.</p> <p>Modification of Symmetry outlier logic.</p>	<p>Additional specialties included.</p> <p>100 new quality and clinical rules added.</p> <p>Addition of confidence intervals around efficiency analysis.</p> <p>Modification of outlier methodology.</p> <p>Minimum number of cases increased.</p> <p>Scorecard enhancements.</p> <p>Expanded number of conditions evaluated for quality.</p>
Use of Resource Use Measures				
Tiering	No. BlueChoice network is a separate health plan product, with a lower premium.	Yes, two-tiered Harvard Independence Plus Plan with \$10 beneficiary co-payment differential.	Yes, two-tiered Aexcel network product with \$10-\$15 beneficiary co-payment differentials.	No, different designations for providers, but no differential co-payment for beneficiaries.
Physician feedback	Group level reports issued (at Tax-ID level).	Confidential reporting of cost areas (pharmacy, outpatient, inpatient) and episode categories driving overall efficiency score.	Confidential reporting of cost areas (pharmacy, outpatient, inpatient) driving high-cost episodes.	<p>Reports mailed to physicians at individual and group level.</p> <p>Physicians also receive log-in information to access detailed data online (at patient level).</p>
Pay-for-performance	No	No	Not in Seattle, but small component of P4P program in Aetna's Northeast plans.	Yes, piloting an automatic fee schedule enhancement.
Consumer reporting	Web-based portal where anyone can view BCBSTX designation on affordability (RACI score) and quality (EBM indicators).	Web-based portal where beneficiaries can view HPHC tier designation.	<p>Web-based portal where beneficiaries can view Aexcel tier designation.</p> <p>Client ability to create custom physician search tool that prioritizes Aexcel physicians.</p>	Web-based portal where anyone can view United star designations.

Table 1 (*continued*)

Aspects of Measuring Resource Use and Quality	BCBS	HPHC (GIC) ^a	Aetna	UnitedHealthcare
Key changes over time	Addition of transparency website in April 2007. Considering establishing a 3-tier product (BlueChoice providers, normal PPO, out of network providers)	Tiering some physicians at the individual instead of at the group level, beginning July 2007.	Introduction of price transparency initiatives.	None

^a Some methodological decisions were made by GIC rather than by HPHC.

^b A “gated” approach refers to an evaluation methodology that requires physicians to meet performance goals in one dimension (e.g. quality) before they are eligible to be evaluated in the next dimension (e.g. efficiency). In other words, physicians must pass through each successive “gate” in the evaluation process. A combination approach evaluates multiple dimensions simultaneously (e.g. quality and efficiency at the same time).

resource measurement might be a consequence of characteristics in the markets we visited that allow providers a relatively high degree of market power in negotiating with health plans that may not exist in other markets. Three of the four markets (Cleveland, Seattle, and Boston) have large physician organizations or hospital-based systems with strong affiliations with physician groups. For example, the Cleveland market is dominated by two large health care systems, with most physician groups aligned with one or the other. While this level of provider negotiation power might not be present in other markets, we also selected these particular markets for our study in part because of health plan recommendations about where their approaches were most highly developed and where they had the most operational experience.

2. Health Plan Measurement Approaches

Three plans we visited use a “gated” approach to physician profiling in their local markets, in which physicians must first meet quality measurement standards and only then are measured on resource use. The gated approach emphasizes quality over efficiency/low cost so that physicians are not rewarded simply for being low cost. One plan measures quality and efficiency of physicians independently and simultaneously, but rates physicians more highly if they meet both efficiency and quality standards.

Quality Measures. Methods for quality measurement varied, but are typically claims-based process of care measures usually drawn from nationally-endorsed, specialty-specific measure sets. Examples typically include measures similar to HEDIS or other measures that gauge a physician’s ability to meet specialty-specific national quality guidelines for rates of preventive services, screenings for chronic illness, or delivery of recommended services for specialty-specific conditions. Structural measures, such as use of electronic medical records, are rarely used and there are only few examples of the use of outcome measures such as complication rates for particular conditions. When nationally-endorsed measures are not available for specific

aspects of care, the plans typically do not provide measurement in that area. However, one plan created its own specialty physician-led committees to develop measures filling gaps in selected specialty areas.

Resource Use Measures. All health plans we interviewed use commercially-developed episode grouper-based tools for resource use measurement. The plans use Episode Treatment Groups (ETGs) developed by Symmetry Health Data Systems or Medstat Episode Groups (MEGs) developed by Thompson Medstat, with one plan using the Cave Grouper, developed by Cave Consulting Group, in combination with ETGs. Another plan also used the Anchor Target Procedure Grouper in combination with ETGs. A plan's choice of grouper tool was often driven by existing ownership or licensure of a particular product, although one plan switched from ETGs to MEGs because it believed that MEGs provided superior case mix adjustment.

Specialties Profiled. The number and range of physician specialties profiled in measurement efforts varies by health plan. Three plans focus on a range of selected (8, 12, and 16) specialties, and one plan includes nearly all specialties that provide direct care to patients. Factors related to choice of specialties for measurement include cost of services, availability of well-established quality metrics, stakeholder/purchaser support for selected specialties, and in some cases the potential for controversy in measuring the cost performance of certain specialties. For example, all plans include most internal medical specialties (e.g., cardiology) and surgical specialties (e.g., general surgery or orthopedics). Two plans focus on specialty care rather than primary care because of the higher costs associated with the former, and two plans exclude oncology while one excludes neonatology, because, in the latter cases, plan representatives did not believe that profiling costs of cancer care or treatment of newborns would be accepted in the market.

Technical Measurement Issues Faced by Health Plans. Health plans reported a number of technical challenges in calculating resource use and quality measures.

Sample size. All plans reported challenges with limited sample sizes at the individual physician level, given their relatively small shares a plan's enrollees represent in a typical physician's practice. Although plans often aggregate analysis up to the physician group level, rather than reporting on individual physicians, the plans were still unable to profile as many as one-third of eligible physicians (that is, those network physicians already meeting specialty type or other inclusion criteria) because of data limitations. When adequate data are unavailable for particular physicians or groups, plans usually place providers in the "low" performing category (see more discussion of uses of measures below).

Lack of standards for resource use. Plans categorize physicians in terms of "high" versus "low" resource use based on relative resource use, rather than objective or national consensus-based absolute standards. Plans have generally chosen arbitrary and varying cut points on which to base relative ranking (percentiles) to distinguish physicians. For example, plans might give a high efficiency ranking to physicians in the top 25 or 50 percent of peers in their specialty group with the lowest resource use.

Identification of individual physicians and physician practices. Plans faced issues in identifying physicians accurately and consistently based on claims identifiers, and building these up to multi-physician practice-based indicators. This challenge was particularly evident in the Boston market because the approach there involves use of multi-payer (health plan) data collected by a major state government purchaser. Appendix C provides more detail on this effort.

Attribution to physician and benchmark comparisons. In developing approaches, all plans confront technical design issues that required decisions on, for example, which types of episodes

are appropriate to attribute to different specialties, and which peer groups individual physicians should be compared to (based on geography and specialty). The design process is largely iterative, starting with a basic approach and then making adjustments/exceptions in specific areas in response to market reaction and/or further data analysis. Health plan staff spend a good deal of time on these issues, “getting down in the weeds” and examining the appropriateness of individual cases when particular types of episodes are attributed to specific physicians, or the appropriateness of comparing certain sub-specialties with one another in selected sub-markets. Exceptions are often made to overall policy decisions given specific physician access concerns, particularly in isolated markets. While the commercial episode grouper software packages plans use have default settings for some of these technical choices related to attribution, there are no national consensus guidelines for making these decisions, producing considerable variation in approaches across plans.

Price-adjusted comparisons. All plans have not adjusted for price differences in comparing the resource use of providers. Health plans generally wanted to take negotiated prices into account when measuring efficiency; therefore providers with higher payment rates would be viewed as more costly, all else being equal. Some plan staff said that it would be useful to generate both price-adjusted and non-price adjusted results in the future.

Uses of Measures in the Marketplace. Health plans visited for this study use quality and resource use measures in four main ways: (1) creation of network “tiers” within an existing HMO or PPO product, (2) development of a lower cost insurance network product built around a smaller “high-performing” network, and/or (3) feedback reports to physicians and (4) “transparency” efforts focused on reporting of performance to consumers and/or purchasers. For tiered network approaches, consumers have copayments at the point of service that are typically \$10 to \$20 lower if they see physicians listed in the higher performing tier of the

network.² For some purchasers or employer groups, plans would use tiers only for transparency efforts, not applying copayment differentials. That is, consumers could be made aware of whether a given physician was in a higher or lower tier, but would not pay a higher or lower copay. In low-cost insurance network products, consumers, and employers if they contribute, pay a lower premium if they purchase a product with a network limited to high performing providers. All four plans implemented some form of transparency/public reporting approach, usually combined with a tiered network or high performing network product approach. Transparency effort usually involved web site reporting to consumers and feedback reports sent to physicians. Among the site visit health plans:

- Two plans (Aetna and Harvard Pilgrim) offer tiered network products, combined with web site public reporting efforts for consumers and feedback reports to physicians. Financial incentives to use providers in one tier versus another are small (for example, a copay differential of \$10 to \$20).
- A third plan (Blue Cross Blue Shield of Texas) offers a separate high performing network product, combined with web site public reporting to consumers and feedback reports to physicians. Enrollment in the high performing network product has been low.
- A fourth plan (United Healthcare) focuses its efforts primarily on public reporting to consumers and feedback to physicians.

Pay-for-performance efforts, using either quality and resource use measures, in which health plan payments to providers are adjusted based on performance were not operational for the plans in the markets we visited, although one plan (Aetna) was in the process of rolling out a pay-for-performance approach in selected markets.

Communication with Physicians. In addition to the technical measurement issues discussed in the section above, health plans devote staff time to consumer and physician

² Health plans typically divide their network into two tiers of providers (high and low performing), although some have three or more tiers.

communication efforts. Consumer targeted reporting of physician performance is provided on each plans' web sites and/or in provider directories, using simple indicators of tier designation—based on underlying scores. However, plans vary in terms of exactly how this information is presented. For example, plans use different symbols such as stars or ribbons, and they varied in terms of whether cost and quality information was presented separately or together. Plans have made modifications to their approaches over time based on market feedback.

Plans also varied in terms of the extent of communication efforts made with physicians both before and during the measurement process and the amount of detail presented to physicians on their scores.. Most plans had relatively little communication prior to sending initial feedback reports or letters, but were in the process of increasing the dissemination of information. For example, one plan notified physicians about their tier designation in a letter without sending any prior information about their measurement efforts. This plan is now working closely with physician representatives to further educate physicians about their measurement activities. Another plan is in the process of disseminating more detailed feedback reports that includes service-specific information related to a particular physician's score, after providing only summary information in a first round of reports. For other plans, more detailed information is only available upon request from physicians. Staff at all of the plans we visited noted that physician communication efforts can be improved and that future investment in enhanced physician feedback methods is important.

3. Physician Reaction

In each market we visited, we talked to representatives from local and/or state medical societies and physicians/administrative staff from both small and large physician practices. In some cases, physicians were invited to the medical society's office where group interviews were held.

Physicians' reactions to health plan measurement efforts were mixed, and appeared to fall into one of three categories. First, many have fairly limited familiarity of specific health plan efforts. But even without such familiarity, many expressed skepticism about the benefits of or need for efforts to measure costs of care.

The second category of physician reaction included fewer providers than the first, consisting of providers who have somewhat greater awareness of specific health plan efforts to measure resource use but who are highly dissatisfied with specific aspects of these measurement efforts. Their dissatisfaction was sometimes associated with their own experiences with having received a low performance rating, having been excluded from a network, or placed in a lower tier based on what they viewed as inaccurate performance measurement. Some physicians who scored highly or who were in the highest tier also raised concerns about the methods and data used in implementing these measures. At the same time, many physicians in this group recognized the importance of resource use measurement in principle, and noted that effective feedback could have effects on physician behavior in constraining resource use while maintaining or improving quality.

Physicians in this second category also raised concerns about the level of communication they had received from health plans. For example, some physicians said they received a letter indicating that they scored lower or higher than a benchmark score on resource use and/or quality measures, but receive little additional information explaining how the scores were developed or why their score deviated from the benchmark. Others said they have learned of their tier designation through a patient or a colleague, rather than through direct communication from the health plan. Finally, some physicians in this category said that their requests for more information are not fulfilled by health plans.

Physicians in a third category, comprising primarily larger multi-specialty or primary care group practices, are also fairly familiar with measurement efforts like the second group, but expressed more positive views about measurement effort and are more active users of the information. We talked to two physician groups in this category. One of these groups was large, as it included more than 200 physicians. The leader of the large group practice said that cost-efficient, high-quality care is core to its mission, and he believes that success in the marketplace is likely to be increasingly based on performance on quality and resource use measures. Thus, this group practice is in the process of investing substantial staff and information technology resources to conduct internal measurement of cost and quality, and is working with health plans to obtain detailed data on the group practice's own performance.

Physicians had specific comments on health plans' resource use and quality measurement efforts. Nearly all physicians interviewed commented on the inconsistency of approaches used by different health plans in the same local market. Even in the Boston market, where the episode grouper tool is applied to multi-payer data, each participating health plan is allowed to use the results in different ways to develop its network tiers (see Appendix C). Physicians said that, consequently, they could be rated highly in one plan network, poorly in another network, or excluded altogether in yet another because their specialty was excluded or because of inadequate data. To the extent they were aware of the measurement techniques themselves, physicians also commented that health plan methods varied widely in terms of the measures selected and how those measures were applied when making network decisions.

Several physicians also had concerns about the validity of the claims data used for measurement purposes, and said that sometimes results did not match their own medical records. These comments were usually aimed at quality measures, rather than resource use measures based on episode groupers.

Although physicians had mixed feelings about being rated by health plans, they generally did not perceive that benefit design and network arrangements based on these measurement efforts were having a large effect on consumer choice of physician. This perception is consistent with health plans' views. Physicians attributed this to the relatively small amount of publicity any one health plan could receive from making physician quality or cost results available, and in the relatively low patient cost-sharing differences among providers placed in different network tiers.

Physicians said that changes in cost or quality were more likely to come from their own internal interest in improving their costs or quality. However, they noted that health plans did not provide much useful detailed information about the measures or explain how the measures were being used to create physician ratings or making network decisions. They also noted that health plans have not involved physicians in the measure development process early enough. At the same time, many physicians—particularly those in smaller practices—do not have time to absorb information on methods even when health plans do provide details.

A related concern expressed by physicians who had received a low rating from a plan was that they did not know why they had received their particular scores, and they did not know what they needed to do improve their scores. Their common reaction to health plans' feedback reporting for efficiency measures was that the information was not “actionable” or service-specific. For example, physicians said they might be told that the costs of particular episodes attributed to them were high compared to their peers, but they were not given information about which services were driving these cost differences (such as more frequent patient admissions, higher rates of ordering tests, or more frequent medication prescriptions). In contrast, physicians said that health plan feedback reporting on most quality “underuse” measures (such as rates of

preventive services) included information that usually made it easier to identify the actions required to improve performance.

A common cross-cutting theme for physician reactions was the desire for standardized measurement, common benchmarks, and coordinated feedback approaches among plans, along with more actionable information on how physicians can improve performance in measured areas. Many physicians noted that aggregated ratings on episode costs should be accompanied with “drill downs,” or supplementary service-specific information, allowing physicians to see how they compare to their peers on delivery of particular services or referral patterns for different types of episodes. Health plan staff interviewed agreed that this was the cutting edge of efforts to develop and use quality and resource use measures. Physicians agreed that if this type of standardized information could be generated (ideally just once in a given period using multi-payer data) and provided in an easy-to-understand and non-punitive fashion, physicians would be more likely to respond with efforts to increase quality while reducing costs in areas where they deviated from their peers.

In summary, physician comments and recommendations for future improvement in quality measurement include:

- ***Standardize Measurement and Feedback Design.*** Future measurement efforts would benefit from standardization of measurement approaches—and ideally, pooling of data across payers—to provide consistent and clearer messages to physicians and consumers about comparative achievement of resource use and quality performance goals. Many physicians commented that Medicare could lead the effort to standardize these approaches.
- ***Use Well-Established and Achievable Resource Use Benchmarks.*** Measurement and physician feedback efforts could also be improved with the development of consensus on physician resource use goals or benchmarks that are consistent across measurement efforts and that are consistent with physicians’ specialty and practice circumstances. In particular, physicians were concerned that resource measures that compare observed to expected (or average) performance was a moving target with no objective standard for efficiency established.

- ***Increase Physician Education and Collaboration.*** Thus far, it appears that physician involvement in quality measure development is more well-established than in development of resource use measures. Our interviews indicate that most physicians are not familiar with, or do not understand well, how resource use measurement tools such as episode groupers work or their potential uses for practice improvement. Future performance measurement efforts can benefit from ongoing education and collaboration with physicians to clarify the goals, methods, and uses of resource use measures.
- ***Produce Actionable Feedback Reports.*** Our study indicates that in order to improve physician response to performance measurement, resource use feedback reports need to provide more detailed information related to service use or referral patterns so that physicians can make more informed decisions to constrain use of resources without sacrificing quality of care.

C. LESSONS FOR FUTURE MEASUREMENT EFFORTS

The private health plans we visited had several years of technical experience implementing physician resource use and quality measures, particularly with the use of episode grouper tools to compare physician costs of care. The Medicare program can draw on their experience for considering potential future uses of these measures in several ways.

Implementing Resource Use and Quality Measures Together. In each of the markets we visited, physician resource use measurement has been implemented in combination with process-of-care quality measurement, usually through a gated approach where physicians who first meet quality standards are then rated on cost. Health plans believe that resource use measures that are not accompanied with quality measures may not be well received by physicians and may remind physicians of “economic profiling” efforts used in prior managed care arrangements that attempted to remove physicians from plan networks based solely on costs. This view appears consistent with the physicians’ own views, although we do not have a direct test of this proposition since all of the physicians we spoke with are measured on both quality and resource use. At the same time, some physicians believe that cost measurement is still the primary focus and motivation behind health plan efforts.

Understanding Physician Reaction to Measurement Efforts. Physician reaction to resource use evaluation has been mixed, ranging from limited awareness and skepticism to familiarity and more active use of measures for internal cost containment and quality improvement initiatives. From this study, it is not possible to identify the factors that determine physician reaction, although larger practices were more likely to be accepting of measures. Representatives of these practices noted that staff had more resources to work with health plans directly and study the information received from plans.

Exploring Multiple Uses of Performance Measures. Health plans visited in this study are in the early stages of experimenting with different uses of performance measures. Efforts to date focus on relatively simple feedback reporting to physicians and on encouraging consumers to use high performing network tiers through public reporting and limited financial incentives for consumers. Health plan staff note that these efforts are likely to evolve over time as they assess the effects of different approaches, including enhancements to physician feedback, public reporting, and pay-for-performance.

The inability to measure a substantial portion of physicians due to small sample sizes has been an important technical limitation to using quality and efficiency measures to initiate broader reform. As a payer with large market share, Medicare could provide larger sample sizes for physician-level measures of performance than most private health plans.

APPENDIX A
INTERVIEW PROTOCOLS

INTERVIEW PROTOCOL FOR HEALTH PLANS

EPISODE GROUPER PRODUCT CHOICE, METHODOLOGY, AND RATIONALE

1. Episode groupers in the context of other resource use or quality measurement activities

- 1.a How long has the plan been measuring resource use?
- 1.b Over time has the health plan refined how it measures resource use?
- 1.c How long has the plan been using episode groupers in this measurement?
- 1.d Does the health plan use per capita measures of resource use?

2. Episode grouper choice

- 2.a Which episode grouper product is the health plan using and why?
- 2.b Has the health plan customized the grouper product at all?
- 2.c Has the health plan ever changed grouper products? If so, why?

3. Episode grouper technical specifications

- 3.a How does the health plan attribute the results of resource use analyses to providers?
 - Does the attribution methodology vary by specialty? By plan model (e.g. HMO vs. PPO)?
 - Does the plan engage in single attribution or multiple attributions?
- 3.b At what level does the plan measure providers (i.e. individual or group level)? Why was this level chosen—for technical reasons? Business reasons?
- 3.c How does the plan determine the minimum number of episodes necessary for resource use analysis to be appropriate for a physician or physician group?
- 3.d Does the plan measure resource use for all providers or just for certain specialties?
 - Primary care vs. specialists?
 - What percentage of physicians in your network do you measure? How is this affected by the minimum number of episodes applied to a physician or group?
- 3.e If certain specialties, why were those specialties selected?
- 3.f Does the plan compare resource use results across specialties for a given condition/episode?
- 3.g Are physician-level results from the groupers stable over time? If the plan uses (or has used) more than one grouper, are results stable across groupers?

- 3.h Does the plan look for and find geographic variation in its episode-grouper driven analyses? Does the plan treat it as a result of the analysis (e.g. follow up with strategies addressing the geographic area), or adjust results for it, or both?
- 3.i Does plan adjust for price differences across providers in calculating efficiency scores?

4. Challenges and lessons with episode groupers

- 4.a What are the most significant technical challenges the health plan has faced with the grouper products that they have used? How have these been resolved?
- 4.b What are the most difficult policy choices (such as which specialties to profile, which thresholds to use, etc.) that the plan has faced in using episode groupers? Why have these been resolved as they have?

USES OF EPISODE GROUPEL-BASED MEASURES

5. Uses of episode grouper-based measures

- 5.a Is the health plan using episode-grouper driven analyses in any of the following ways: confidential feedback, tiered plans, recognition programs, narrow networks, pay-for-performance?
- 5.b Does the health plan make the results of its episode-grouper driven analyses public (to other providers, to the general public, or to its members)? Why or why not? In what degree of detail are the results presented?
- 5.c Over time, has the plan changed how the results of its analyses are used and shared? If so, why?
- 5.d How does the plan incorporate the results of episode grouper analyses into payment policy? Has this changed over time and, if so, why?
- 5.e Do you have plans to use resource use measures differently in the future? If so, what are they?

INTERACTION OF EPISODE GROUPEL-DRIVEN ANALYSES WITH OTHER QUALITY MEASUREMENT ACTIVITIES

6. Interaction of quality measurement and resource use management

- 6.a For those products and providers where resource use is measured, is quality also measured? If not, is this desired or planned? Why/why not?

- 6.b Does the health plan use a software package to monitor quality in conjunction with its episode grouper analysis? If so, what product does the health plan use and how does it interact with the episode grouper software?
- 6.c Has the health plan observed any correlation between resource use efficiency and quality?
- 6.d Do plan staff believe quality has been influenced by use of the groupers (e.g. identification of underuse, overuse or misuse of services)? How so and why?
- 6.e What is the relative importance of resource use measurement and quality measurement where both are measured?
 - Are both assessed concurrently (i.e. mixed quality/efficiency score) or is there a “gated” approach (i.e. physician groups must be efficient before they are evaluated for quality, or vice versa)?
 - Why have these decisions been resolved the way that they have?

INTERACTIONS WITH AND REACTIONS FROM THE PHYSICIAN COMMUNITY; HOW PHYSICIANS RECEIVE AND CAN USE EPISODE-GROUPER DRIVEN DATA

7. Interactions with the physician community

- 7.a Were physicians involved in the design of the episode-based analyses?
- 7.b How does the health plan communicate the results of episode analyses to physicians? Were other options considered, and, if so, why were they rejected?
- 7.c At what level of detail are physicians given data? Why was this level of detail selected? [Can we get a blinded copy of a typical report?]
- 7.d How are physician groups using the data that have been provided by the health plan?
- 7.e To what degree has the health plan interacted with physician associations versus individual physicians or physician groups? Are there certain physicians known to be most influential, and have you made any special effort with them in the process?
- 7.f Please describe any mechanisms that exist for physicians to appeal their ranking/placement following episode-grouper analyses.
 - How many/what percentage of the health plans’ physicians have used this appeal mechanism? Are they generally successful?

8. Reactions from the physician community

- 8.a How and to what extent have physicians reacted to the use of episode-based analysis?
- 8.b To what extent do physicians view episode-based analyses as credible? Why?

- 8.c What appears to affect physician acceptance of episode grouper analysis? For example:
- How well physicians perform
 - What episode group analysis is used for
 - How the results are presented
- 8.d What characteristics of different physician groups or regions do you believe have influenced the types of responses you have received from the physician community?
- 8.e How has the proportion of business that a health plan represents for a given physician affected responsiveness to being profiled?
- 8.f Have other resource measurement activities in the market affected physician response to measurement efforts? If so, how?
- 8.g Have any other factors influenced physicians' reaction to resource use measurement?

IMPACT AND LESSONS LEARNED

- 9.a What have been the primary benefits of using episode-grouper driven data? The primary challenges?
- 9.b What are the most important factors affecting physician response to the episode-grouper driven products?
- 9.c Have cost savings been achieved through resource use measurement activities? What role have episode-grouper driven analyses played in these savings? Through what mechanism have these savings primarily occurred?

PROTOCOL FOR MEDICAL ASSOCIATIONS

I. MARKET BACKGROUND AND HEALTH PLAN INTERACTIONS WITH PHYSICIANS

1. Can you provide an overview of how the health plans in your area have been using episode groupers or other measures of physician resource use?
2. Are most health plans in your market measuring physician resource use?
3. How have health plans involved physicians in developing or revising episode-based analyses?
4. To what extent have the health plans interacted with your medical association about episode-grouper based analyses?

II. PHYSICIAN USE OF DATA

1. How are physician groups using the data that have been provided by the health plans?
2. What types of infrastructure at the physician group or individual practice level do you believe are necessary to interpret and respond to episode-grouper driven data? To what extent do you believe these resources are currently available to physicians?
3. Do physician groups believe that quality has been influenced by the use of groupers (e.g., identify overuse, underuse, or misuse of services)?
4. What have been the most significant effects of private health plans' use of episode-grouper-based analyses?

III. PHYSICIAN REACTION TO PROFILING

1. Does the medical association have any official policy on episode-grouper based analyses or related measurement efforts? If so, please describe the policy, how it was developed, and any related initiatives.
2. Please describe physicians' reactions to the episode-grouper based analyses that health plans are performing in this market.
3. To what extent do physicians view episode-based analyses as credible for measuring efficiency or resource use? Why?
4. If multiple health plans in your market are using episode-grouper based analyses, which analyses seem most credible? What characteristics (either about the methodology itself, or the relationship of local physicians with the health plan) make these approaches most credible?
5. Do health plans generally make the results of episode-grouper based analyses publicly available?

- If so, at what level of detail are the results released to the public?
 - How does the medical association view this practice?
6. Have health plans in your market have incorporated quality measurement with their episode-grouper based analyses? If so, how?
 7. What does the medical association believe is the appropriate balance between quality measurement and resource-use measurement? Are health plans achieving this balance?
 8. What have been the most important influences on physician reaction to episode-grouper analyses?

PROTOCOL FOR PHYSICIAN GROUPS

I. MARKET BACKGROUND AND HEALTH PLAN INTERACTIONS WITH PHYSICIANS

1. Can you provide an overview of how the health plans in your area have been using episode groupers or other measures of physician resource use?
2. How many health plans in your market are using episode groupers or other methods to measure resource use?
3. How have health plans involved you or your practice in developing or revising episode-based analyses?
4. To what degree have the health plans interacted with physician associations versus individual physicians or physician groups about episode-grouper based analyses?
5. For how many years has your practice been the subject of resource use profiling?
6. [If applicable] Do you use resource use measurement data to assess physician performance within your own practice?

II. PHYSICIAN USE OF DATA

1. How do health plans communicate the results of episode-grouper based analyses to your practice?
 - At what level of detail do physicians receive data from the health plan? Is data at this level of detail useful for you?
2. Are you aware of how health plans rank your resource use relative to other physician groups?
 - Are there any areas where the data seem to show you as particularly high or low? If so, did you feel the need to investigate why?
 - [If no investigation:] Do you know or have a theory about why?
 - [If investigation:] What did you find?
3. What do you do with the resource use data plans send? Has the data ever led you to make a change in practice? If so, how? If not, why not?
4. What types of infrastructure at the physician group or individual practice level do you believe are necessary to understand and use episode-grouper driven data? To what extent do you believe these resources are currently available at your practice? To physicians more generally?
5. What degree of support, if any, have the health plans provided to your practice to help you interpret and respond to episode-grouper based data?

6. Does your physician group believe that quality has been influenced by the use of episode groupers (e.g. identification of service underuse, overuse, or misuse)?
7. What have been the most significant impacts on your practice from private health plans' use of episode-grouper-based analyses?

III. PHYSICIAN REACTION TO PROFILING

1. Please describe your reaction to the episode-grouper based analyses that health plans are performing in this market, and, to the extent you can, the reaction of physicians more broadly.
2. To what extent do you view episode-based analyses as a valid method for identifying efficient or inefficient practices? Why?
3. If you are profiled by multiple health plans, have you looked at how consistent they are?
 - Which analyses seem most valid?
 - What characteristics (either about the methodology itself, your practice, or your relationship with the health plan) make some approaches better than others?
4. Do health plans that you have contracts with make the results of episode-grouper based analyses publicly available?
 - At what level of detail are the results released to the public?
 - How do physicians view this practice?
5. Do health plans incorporate quality measurement with episode-grouper based analyses of your practice?
6. What do you believe is the appropriate balance between quality measurement and resource-use measurement (assuming some level of each)? How close has the health plan come to this balance in your estimation?
7. (If applicable.) Are you aware of any mechanisms through which you can appeal your placement or ranking that was derived from episode-grouper based analyses? What is your perception of these appeal mechanisms?
8. Are you aware of any policy opinions that your state or local medical association has expressed about the use of episode groupers? Do you concur with these positions?
9. What are the most important influences on your reaction to episode-grouper analyses?

APPENDIX B

SEATTLE SITE SUMMARY

Contract No.: RFP03-06-MedPAC/E4016631
MPR Reference No.: 6355-300

MATHEMATICA
Policy Research, Inc.

Site Visit Summary: Seattle Washington

July 2, 2007

*Margaret Colby
Timothy Lake*

Submitted to:

Medicare Payment Advisory Commission
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Project Officer:
Niall Brennan

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Timothy Lake

CONTENTS

	Page
SUMMARY	B.1
AETNA’S MEASURES OF PHYSICIAN RESOURCE USE	B.2
Motivations for Physician Resource Use Measurement	B.2
Selection of Physicians for Resource Use Measurement.....	B.3
Methodological Considerations in Resource Use Measurement.....	B.3
USES OF EFFICIENCY MEASURES BY HEALTH PLANS	B.4
Tiered Networks.....	B.4
Confidential Feedback Reports	B.5
Reporting to Consumers.....	B.5
Pay for Performance.....	B.5
Approaches of other Seattle-area Health Plans	B.6
PHYSICIAN REACTION	B.6
Interactions between Aetna and Physician Groups	B.7
Physician Reactions to Resource Use Measurement.....	B.7
Physician Reactions to Health Plans’ Uses of Efficiency Measures.....	B.8
Physician Uses of Efficiency Measurement.....	B.8
KEY LESSONS AND CONCLUSIONS.....	B.9
Effects of Efficiency Measurement on Resource Use.....	B.9
Key Benefits and Challenges	B.9
Assessment of Future Market Trends	B.9

SITE VISIT SUMMARY SEATTLE, WASHINGTON

This report summarizes information obtained in interviews with health plan officials, physicians, and health care purchasers that were conducted during a site visit to Seattle, Washington, from June 11-12, 2007 to explore health plans' use of episode groupers for efficiency measurement. Following a semi-structured interview guide, MedPAC and Mathematica Policy Research, Inc. staff spoke with Aetna executives, the Washington State Medical Association (WSMA), the Puget Sound Health Alliance (PSHA), and three area physician groups, two of whom were large multi-specialty clinics and the third was a smaller primary care-oriented group. The purpose of the interviews was to learn about Aetna's technical experiences using episode grouper software to measure physician efficiency, as well as to obtain information about physicians' reactions to private health plans' use of this type of measurement.

SUMMARY

- In response to requests from its clients, Aetna developed a tiered network product using episode-based efficiency analyses. The product separates selected specialists who account for the majority of health care costs in the Seattle market into two tiers based on their quality and efficiency performance.
- Physician reaction to efficiency measurement ranged from acceptance to skepticism. Some multi-specialty physician groups are using episode-based analyses to identify opportunities for efficiency improvement. Others believe that episode-based measurement is valid in principle, but difficult and resource-intensive to implement in a credible way. WSMA has several concerns about efficiency measurement and would prefer to focus on developing robust quality measures.
- Physicians' concerns about the use of episode-based efficiency measurement fall into two broad categories: concerns about the variability of their performance scores across different health plans, and concerns about the actionability of data reports received from health plans.
- Health plans' recommendations for action regarding the use of specific services may be more important for small physician groups, who lack the internal resources to analyze episode-based efficiency data themselves. However, it is not clear whether more specific recommendations for efficiency improvements would be well-received or viewed as overly prescriptive.
- The PSHA, a multi-stakeholder initiative that includes employers, health plans, and physician groups, is exploring long-term plans to aggregate data across health plans and perform episode-based efficiency analyses. Several stakeholders were optimistic that this process would resolve issues of small sample size and help standardize plans' methodologies, but others noted the challenges of the Alliance in balancing purchaser, plan, and provider concerns in this area.

- In the short term, health plans in the Seattle area will continue to explore the use of efficiency measurement for pay-for-performance (P4P) incentive systems and consumer reporting initiatives. Support from Seattle-area employers will remain an important motivator for health plans' continued efforts at efficiency measurement, and financial incentives may be an important element for motivating behavioral changes within physician groups.

BACKGROUND ON HEALTHCARE MARKET IN SEATTLE

Several large Seattle employers, including Boeing, Starbucks, REI, and Costco are influential in health plans' benefit designs, and have continued their historic support for both quality and efficiency measurement initiatives. For example, in response to a specific request from Boeing, Regence BlueShield used episode-based efficiency measurement to develop its Regence Select Network, which was introduced in 2006 and subsequently dissolved following complaints from beneficiaries and legal challenges from providers. In addition to their influence on plan-specific initiatives, Seattle-area employers have joined together with health plans and physician groups in a multi-stakeholder initiative called the Puget Sound Health Alliance, which is exploring long-term plans to aggregate data across health plans for quality and efficiency analyses, despite the reaction from providers and beneficiaries to the 2006 Regence product.

Overall, Premera Blue Cross and Regence BlueShield are the dominant insurers in the Seattle area, with Aetna holding approximately 8-10% of the private health insurance market. Group Health of Puget Sound is a fourth major participant in the Seattle market and uses a predominantly staff-based model, rather than a community network of providers; however, overall managed care penetration in the Seattle area was only 16% in 2005, nearly half of the national HMO penetration rate among large metropolitan areas.³ While some physician groups include more than 300 providers and have significant negotiating power with health insurers, smaller groups of providers are also common. In general, physician groups tend to refer patients for inpatient treatment to Swedish Medical Center or Virginia Mason Medical Center, as they are the two dominant hospital systems in the Seattle market.

AETNA'S MEASURES OF PHYSICIAN RESOURCE USE

Motivations for Physician Resource Use Measurement

Aetna began measuring physician efficiency in 2002 in order to address concerns that its clients had about the quality and rising costs of health care. The plan has developed physician efficiency measurement by analyzing claims data with episode grouper software tools. In conjunction with quality measurement, these data are used to create a tiered network product called Aexcel, in which beneficiaries are steered towards Aexcel-designated physicians via co-payment differentials. Although Aetna holds only 8-10% of the private health insurance market in Seattle, Aetna has been able to implement this model due to support from major employers in

³ Center for Studying Health System Change. "Community Quality Efforts Expand as Seattle Health Plan Products Evolve. Community Report: Seattle, Washington." September 2005.

Seattle, such as Starbucks, REI, and Costco. Aetna indicated that these clients view physician efficiency measurement as a long-term investment.

Selection of Physicians for Resource Use Measurement

Aetna has chosen to focus on selected physician specialties that account for the majority of health care costs. In 2002, Aetna began profiling six specialties and expanded to twelve specialties the following year. These specialists account for 56 percent of Aetna's health care costs in the Seattle market. Aetna has excluded some specialties from the Aexcel-designation process for a number of reasons, including a lack of valid quality measures and concerns about beneficiary reaction. For example, oncologists have been excluded because of the emotional nature of the care these providers deliver and the lack of clear evidence-based guidelines for the specialty. Aetna also has chosen not to measure the efficiency of primary care physicians (PCPs) because patients tend to have more established relationships with these providers. Additionally, Regence BlueShield's prior efforts at profiling PCPs in the Seattle area had generated significant resistance from beneficiaries. Aetna was also concerned that disrupting established 'medical home' relationships with PCPs might increase costs.

Methodological Considerations in Resource Use Measurement

Aetna has periodically refined its efficiency measurement methods since the first round of analysis in 2002. Aetna began processing claims data using Symmetry Episode Treatment Groupers (ETGs), selected because the company had experience using that software product for predictive modeling and actuarial purposes. Early experience demonstrated that Symmetry ETG-based efficiency scores were susceptible to variations in patient case-mix; a physician might appear more or less efficient based on the severity of illness of his patients, rather than based on the physician's own care patterns. Aetna was concerned that this variability would cause the Aexcel-designation turnover rate for physicians (movement from one tier to another) from year to year to be unacceptably high.

To create more stability in the Aexcel-designation process, Aetna began analyzing claims with both the Symmetry and the Cave Grouper in 2005; the Cave Grouper uses a different methodology to adjust for case severity. Physicians who rank in the top 50 percent under both episode groupers are Aexcel-designated. Aetna has not had enough experience with this combined Symmetry/Cave approach to draw conclusions about the stability of its results; however, it believes that a turnover rate of 20 percent may be appropriate, given practice pattern changes and contractual modifications that occur over time.⁴ Overall, Aetna has found it challenging to evaluate Aexcel network performance over its five-year history of operation because Aetna has changed its efficiency measurement methodology somewhat each year.

⁴ Aetna wants its efficiency scores to capture differences in contractual payment rate agreements; therefore, they have not explored making price-neutral adjustments (use of standardized payment rates) to their analyses.

USES OF EFFICIENCY MEASURES BY HEALTH PLANS

Health plans may use efficiency measures in several ways, including the production of confidential physician feedback reports, public reporting to consumers, P4P initiatives, creating tiered networks, and creating more narrow or selective networks. Aetna has primarily applied its efficiency measures to create a tiered physician network, although it is exploring other applications in Seattle and other markets.

Tiered Networks

Construction of the Network

Aetna's primary use of efficiency measures has been to develop a two-tiered network product called Aexcel. There are three elements that determine whether a physician receives Aexcel designation, which means they are included in the first, or top, tier:

- First, physicians must have at least 20 episodes of care in order to be eligible for the first tier. In the Western Washington region, which includes Seattle, about 20-25 percent of physicians have fewer than the minimum number of episodes and are placed in the second tier.
- Second, each specialist is assessed on quality measures that are well-accepted and endorsed by his/her specialty profession. Only 5 percent of physicians are eliminated from the first tier at this stage, due to sensitivity about implicitly labeling a large number of non-designated physicians as "low-quality."
- Third, the remaining physicians are evaluated on efficiency, and the top 50 percent most efficient physicians, as measured by both the Symmetry and Cave ETGs, receive the Aexcel-designation and are placed in the first tier.

Aetna aggregates data to the specialty group practice level for the Aexcel-designation process. It has decided not to tier at the individual physician level, due to concerns about beneficiary complaints. For example, beneficiaries would likely complain if they faced barriers to seeing another cardiologist within the same practice when their usual cardiologist is unavailable. Although specialty groups are tiered as a unit, Aetna has Aexcel-designated some specialties and not others within large multi-specialty groups. Contractual negotiating power with large provider groups, specific plan sponsor preferences, and geographic access considerations also influence Aexcel-designation decisions.

- ***Network Incentives***

In general, the Aexcel product steers beneficiaries towards Aexcel-designated physicians through co-payment differentials; no providers are explicitly excluded from the network. Because Aetna offers this product only to its self-insured clients, the co-payment differentials are specific to each plan sponsor.

Most clients currently have modest co-payment differentials of \$10-\$15 between Aexcel-designated physicians and other Aetna physicians. The evidence Aetna has on the importance of co-pay differentials for steering beneficiaries is mixed. One Aexcel client has not introduced any differentials and has found it difficult to gain steerage and realize savings. Another client has engaged in an extensive information campaign, and almost three-quarters of its beneficiaries now visit Aexcel-designated physicians, although they have no financial incentive to do so. Among clients with co-payment differentials, Aetna has not formally evaluated the impact of financial incentives on beneficiary choices, but believes that such tools help increase steerage.

Confidential Feedback Reports

Aetna provides its physicians with reports that indicate the episodes for which they are more costly than their peers and show a breakout of the medical cost categories, (such as inpatient, outpatient, and pharmacy costs) that are driving the physicians' overall scores. Aetna has the capacity to drill down to the individual line-item level if physicians request additional data to help them understand their efficiency scores. Staff indicated that an initial two-to-three year confidential reporting period before creating tiers might have been helpful for eliciting physician feedback and making modifications to the measurement process.

Reporting to Consumers

Aetna beneficiaries are able to view Aexcel-designations through a web-based interface, in which Aexcel-designated physicians are indicated with a blue star. For all Aetna providers (those with and without Aexcel-designations), beneficiaries can click on a tab for performance information. The tab briefly describes Aetna's episode volume, quality, and efficiency measurements, and indicates with a checkmark whether the provider meets each standard. Aetna's clients are also able to create customizable physician search tools for their beneficiaries that prioritize Aexcel-designated physicians.

Pay for Performance

Although efficiency measurement has not been actively used in the Seattle area as part of Aetna's P4P programs, the company has begun incorporating episode-based efficiency analysis in its P4P programs in the Northeast. The Northeast program has three modules: quality, efficiency, and patient satisfaction. The efficiency component includes seven to eight measures, and one of these measures is developed from an episode-based efficiency score. Overall, episode-based analysis represents less than 40 percent of the efficiency component, and no more than 10 percent of the overall P4P score.

Approaches of Other Seattle-Area Health Plans⁵

Two other major health plans in Seattle, Premera BlueCross and Regence BlueShield, are engaged in efficiency measurement but do not currently emphasize a tiering approach.^{6,7} Regence BlueShield does not currently emphasize its episode-based efficiency analyses in benefit or network design. However, it attempted to implement a narrow network, called the Regence Select Network, for Boeing employees in 2006. To develop that product, Regence profiled physician groups, including PCPs, using episode-based efficiency analyses. The least efficient groups were excluded from the network and beneficiaries treated by these providers were not covered by Regence. Despite Regence's emphasis on efficiency measurement, letters to affected beneficiaries in the plan implied that providers had been excluded because of poor quality performance. These statements led to the filing of a libel lawsuit and subsequently the dissolution of the Select Network product. In comparison to this effort, Aetna's Aexcel product has been less controversial: PCPs are not profiled, no providers are excluded from the network, and the overall impact for physicians is smaller, given Aetna's smaller market share.

Premera BlueCross currently uses its episode-based efficiency analyses in a P4P incentive system, much like Aetna's program in the Northeast. Premera also has the capacity to provide detailed feedback reporting to providers upon request, has established data transfer relationships with some larger physician groups, and works actively with those providers to identify opportunities for improvement that are highlighted by the episode-based analyses. Given its ability to provide data on a significant proportion of physicians' patients, Premera may be able to provide more robust physician-level measures than Aetna.

PHYSICIAN REACTION

Physician reaction to efficiency measurement ranged from acceptance to skepticism. Some multi-specialty physician groups are using episode-based analyses to identify opportunities for efficiency improvement. Others believe that episode-based measurement is valid in principle, but difficult and resource-intensive to implement in a credible way. Overall, physicians' concerns about the use of episode-based measurement fall into two broad categories: concerns about the variability of their performance scores across health plans, and concerns about the actionability of data reports received from health plans.

⁵ Information about efficiency measurement efforts by Premera BlueCross and Regence BlueShield was obtained through the Seattle-area physician interviews, as well as news articles and a telephone interview with Premera BlueCross.

⁶ "Tiering" refers to the practice of treating subsets or "tiers" of a provider network differently, usually in terms of the co-payments or co-insurance that patients pay out-of-pocket for seeing providers in a particular tier. For example, in Aetna's Aexcel product, patients pay \$10-\$15 more to see a physician who has not earned the top-tier Aexcel designation.

⁷ A fourth major health plan in the Seattle market is Group Health of Puget Sound, which uses a predominantly group/staff model network arrangement rather than a community-based provider network.

Interactions between Aetna and Physician Groups

Aetna has engaged in some proactive communication efforts to solicit physicians' opinions on product design. For example, in the early 2000s Aetna was a participant in meetings that convened several large physician groups and the major health plans operating in the Seattle market to try to standardize plans' episode-based efficiency measurement methodologies. Also, prior to introducing the Aexcel network product, Aetna staff communicated with several of the larger physician groups in the Seattle area. More recently, Aetna has worked with some physician groups to help them improve their efficiency or to reduce resource use. Aetna has also discussed its future plans and methodological approach with the Washington State Medical Association (WSMA), and has utilized WSMA as an intermediary during negotiations with physician groups who are at risk of losing their Aexcel-designation status.

Physician Reactions to Resource Use Measurement

In the Seattle area, physicians' reactions to the concept of episode-based efficiency measurement ranged from acceptance to skepticism. At one end of the spectrum, a large multi-specialty group of about 250 physicians uses episode-based efficiency analyses obtained from Premier BlueCross to identify services where higher-than-network-average costs indicate the potential for savings. The group has incorporated episode-based efficiency analyses as a key element of its business model, anticipating that high performance networks will be an increasingly important aspect of the health care market, and its own success will be tied to the ability to qualify for these network designations. In contrast, another large multi-specialty clinic believes that episode-based efficiency measurement is valid in theory, but difficult and resource-intensive to implement in a credible way. This group prefers to devote its internal resources to quality improvement efforts, and other, less analytically intense programs, like monitoring generic prescribing. Representatives from a small physician group that focuses on primary care were skeptical that episode-based analyses would be a valid mechanism for evaluating PCPs, because they believe a large percentage of their day-to-day activities are not captured by the tool.

Although WSMA believes that physicians would be receptive to changing their practice patterns if credible data about cost-effectiveness were available, it has some concerns about episode-based efficiency analysis. In particular, WSMA is concerned that episode-based analyses rely on claims data that do not adequately incorporate information about care outcomes. An official also explained that a lack of widespread electronic medical records makes it very difficult for physicians to validate or verify the accuracy of efficiency measures that health plans generate. This problem is particularly acute for smaller practices, and WSMA believes that formal processes need to be developed to facilitate discussions between physicians and health plans when there are disagreements about data. Their members have also expressed frustration with the relative ranking system of efficiency measurements, and believe that attainment of cost-effectiveness benchmarks would be preferable. Because of these concerns with episode-based efficiency measurement, WSMA would prefer to focus on developing robust quality measurements.

Physician Reactions to Health Plans' Uses of Efficiency Measures

In the Seattle market, physician groups expressed several concerns with health plans' uses of episode-based efficiency measurement. Their concerns fall into two broad categories: concerns about the variability of performance scores across health plans, and concerns about the data reports received from the health plans. First, some physicians stated that they were concerned about the validity of the data and/or methodology used by each health plan in episode-based analyses because their efficiency scores had fluctuated across health plans when multiple plans (Aetna, Regence, or Premera) had at one time been feeding back data to the groups. For example, one group reported having physicians who were ranked as 50 percent more efficient than their peers by one plan, and 50 percent less efficient by another. In general, physicians felt that this variability undermined the credibility of the measurement process, as well as tiered networks that had been based on the data. They were also skeptical of reports and tiering decisions that had been based on a small percentage of their overall practice. The PSHA, a multi-stakeholder initiative including major Seattle-area employers, health plans, and physician groups, is exploring ways to aggregate data across health plans and perform episode-based analyses. Several stakeholders were optimistic that this consolidated evaluation process would resolve issues of small sample size and help standardize health plan methodologies, though they also noted the challenges the Alliance faces in balancing diverse concerns of different stakeholders in this effort.

The second cluster of physicians' concerns was that the data reports they receive from health plans do not provide data at an actionable level. One group suggested that in order to achieve behavioral changes, the reports need to include two to three concrete steps that a physician or group can take to improve efficiency for a particular type of episode. At the same time, a large multi-specialty group that conducts its own internal efficiency analysis suggested that physicians might perceive health plans' recommendations of specific action items as too heavy-handed, indicating the need for balance and flexibility in health plan communications to physicians in this area. Overall, physician representatives agreed that health plans' synthesis of episode-based data and recommendations for action may be more important for small physician groups who lack the internal resources to perform the analyses themselves.

Physician Uses of Efficiency Measurement

Some large multi-specialty physician groups in Seattle are using health plans' resource use measures, or perform their own internal episode-based analyses with the health plans' data, in an attempt to change practice patterns and become more efficient. For example, one group noted that imaging was making its cost of care appear higher than average. To reduce imaging utilization, the group developed an evidence-based checklist that its physicians use before ordering an image. Other physician groups in the Seattle area that we spoke with have generally not used episode-based analyses or responded to their efficiency scores. One physician group suggested that, given the internal resource requirements necessary to use episode-based analyses in a meaningful way, the typical physician group does not have sufficient financial incentives to use the measurements. Aetna staff confirmed that many physicians do not request more information about their efficiency scores until a patient inquires about their Aexcel designation.

KEY LESSONS AND CONCLUSIONS

Effects of Efficiency Measurement on Resource Use

Based on the set of interviews, it appears that some large providers in the Seattle area are actively using episode-based analyses to identify opportunities for efficiency improvement and to change practice patterns; however, most physician groups have not responded to the tiered networks or incentive programs designed around these measurements. Overall, Aetna anticipates savings of two to four percent in its Aexcel-network plan relative to its other products. Staff noted that it was difficult to estimate the factors driving those savings, since the populations that are covered under each type of plan may be different.

Key Benefits and Challenges

The Seattle market illustrates several benefits and challenges of episode-based resource use measurement. One key benefit is that some large physician organizations are actively using episode-based analyses to improve the efficiency of their practice patterns, which suggests the tool's potential for highlighting areas for resource use improvement. Additionally, early efforts at consumer reporting illustrate efficiency scores' potential to engage beneficiaries as thoughtful health care consumers. Aetna staff also indicated that episode-based efficiency analyses have drawn physician attention to the cost of care, serving as a useful balance, while the health care community in general focuses on quality measurement.

Key challenges include the variability of methodologies and physician scores across plans, which hurt the credibility of episode-based efficiency measurement. Additionally, many providers lack the internal resources to analyze health plan data and develop actionable tasks based on that data that would improve their efficiency. To date, health plans have provided relatively high-level information to physicians. However, it is not clear whether more specific recommendations for efficiency improvements would be well-received or viewed as overly prescriptive. Longer-term political/economic challenges lie in determining the appropriate type and level of financial incentives needed to encourage providers to be more efficient, and to encourage consumers to choose more efficient providers. Employers are generally focused on the use of efficiency measurement to achieve cost savings; however, both WSMA and PHSA have advocated gain-sharing or budget-neutral models that would reward efficient performers from a pool of reserved funds.

Assessment of Future Market Trends

PSHA is assessing ways to coordinate an efficiency measurement initiative with aggregated data from the major health plans. Several stakeholders are optimistic that this collaboration will help standardize methodologies across plans and reduce the administrative burden associated with multiple reporting systems. In the short term, health plans will continue to explore the use of efficiency measurement for P4P incentive systems and consumer reporting initiatives. Support from Seattle-area employers will remain an important motivator for health plans' continued efforts to measure provider efficiency.

SITE VISIT SUMMARY TABLE
AETNA
SEATTLE, WA

Aspects of Measuring Resource Use and Quality		Aetna
OVERVIEW		
Year began using physician resource use measures		2002
Episode grouper or other measurement product(s) used		Symmetry ETG, Cave Grouper
Quality measures used in conjunction with resource use measures		Eight quality measures per specialty that have been endorsed by their professions
TECHNICAL SPECIFICATIONS		
“Gated” approach vs. combination of quality, resource use measures, and/or sample size		“Gated” approach—1) volume of episodes, 2) quality, and then 3) efficiency
Individual provider and/or group level measurement		Individual provider-level measurement, rolled up to group-level for tier designation
Specialties profiled		12: Cardiology, cardiothoracic surgery, gastroenterology, general surgery, neurology, neurosurgery, obstetrics and gynecology, orthopedics, otolaryngology/ETN, plastic surgery, urology, and vascular surgery
Attribution method		Symmetry ETG: Standard, each episode attributed to one physician, to surgeon if a surgery occurs Cave Grouper: Episodes attributed to each physician with more than 20% of costs, possible for single episode to be attributed to multiple physicians
Minimum number of episodes required for profiling a physician		20 episodes
Percentage of potentially eligible physicians profiled		70-75%
Addressing outliers		Modification of Symmetry ETG to eliminate most volatile episodes
Methods for benchmarking/ranking		Quality—top 95% Efficiency—top 50% as designated by both Symmetry and Cave ETGs Western Region is comparison group for Symmetry ETG and Seattle service area is comparison group for the Cave Grouper
Price standardization methods		No price standardization
Data source and/or aggregation (including number of years of data)		Two years self-insured commercial plan data
Key modifications over time		Addition of Cave grouper in 2005, modification of Symmetry outlier logic
USE OF RESOURCE USE MEASURES		
Creating tiered networks		Yes, Aexcel network product with co-payment differentials
Physician feedback		Yes, confidential reporting of cost areas (pharmacy, outpatient, inpatient) driving high-cost episodes

Aspects of Measuring Resource Use and Quality	Aetna
Pay-for-performance	Not in Seattle, but small component of P4P program in Aetna's Northeast plans.
Consumer reporting	Yes, web-based portal where beneficiaries can view Aexcel-designation. Client ability to create custom physician search tool that prioritizes Aexcel physicians.
Key changes over time	Will be introducing price transparency initiatives

APPENDIX C

BOSTON SITE SUMMARY

Contract No.: RFP03-06-MedPAC/E4016631
MPR Reference No.: 6355-300

MATHEMATICA
Policy Research, Inc.

**Site Visit Summary:
Boston Massachusetts**

July 23, 2007

*Margaret Colby
Timothy Lake*

Submitted to:

Medicare Payment Advisory Commission
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Project Officer:
Niall Brennan

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Timothy Lake

CONTENTS

	Page
SUMMARY	C.1
BACKGROUND ON HEALTHCARE MARKET IN BOSTON	C.2
GIC'S MEASURES OF PHYSICIAN RESOURCE USE	C.3
Motivations for Physician Resource Use Measurement	C.3
Methodological Considerations in Resource Use Measurement.....	C.3
USES OF EFFICIENCY MEASURES BY HEALTH PLANS	C.4
Tiered Networks.....	C.5
Confidential Feedback Reports	C.6
Consumer Reporting	C.7
PHYSICIAN REACTION	C.7
Interactions with Physician Groups.....	C.7
Physician Reactions to Efficiency Measurement.....	C.7
Physician Reactions to Health Plans' Uses of Efficiency Measures.....	C.7
Physician Uses of Efficiency Measurement.....	C.8
KEY LESSONS AND CONCLUSIONS.....	C.9
Effects of Efficiency Measurement on Resource Use.....	C.9
Key Benefits and Challenges	C.9
Assessment of Future Market Trends	C.10

SITE VISIT SUMMARY BOSTON, MASSACHUSETTS

This report summarizes information obtained in interviews with health plan officials, physicians, and health care purchasers that were conducted during a site visit to Boston, Massachusetts, from June 18-19, 2007. Following a semi-structured interview guide, MedPAC and Mathematica Policy Research, Inc. (MPR) staff spoke with Harvard Pilgrim Health Care (HPHC) executives, the Group Insurance Commission (GIC), the Massachusetts Medical Society (MMS), and representatives from seven area physician groups (five specialty groups and two primary care practices). The purpose of the interviews was to learn about GIC's and HPHC's technical experiences using episode grouper software to measure physician efficiency, as well as to obtain information about physicians' reactions to private health plans' use of this type of measurement tool.

SUMMARY

- GIC, which self-administers health insurance for employees of the Commonwealth of Massachusetts, is the primary driver of episode-based efficiency measurement of physician practices in the Boston area. GIC views efficiency measurement as a tool for reducing the Commonwealth's health care costs.⁸
- Since 2004, GIC has required its six health plans to submit claims data for their full books of business, which GIC then compiles into an all-payer dataset and analyzes using episode grouper software. Health plans must use the episode-based results and quality metrics to construct tiered physician networks for GIC beneficiaries. Because GIC has not been prescriptive about plans' tiering methodology, the plans have developed several different models.⁹
- Physicians' concerns about GIC's use of efficiency measures fall into three categories: 1) concerns about diverse tiering designations across plans, 2) concerns that episode-based data are not being used constructively to help improve physician practice patterns, and 3) concerns about beneficiaries' reactions to physician tiering.
- Although physicians were very concerned when GIC first announced its physician profiling efforts in 2004, GIC stated that its health plans have received fewer complaints in recent years. Physicians confirmed the decline in their attention to this issue and attributed their response to the minimal impact GIC has actually had on their patient volume or finances to date. While the response from most physicians has

⁸ The Massachusetts Health Quality Partnership has also been testing efficiency measures calculated from multi-commercial payer claims data, but has only used the results for research purposes to date.

⁹ "Tiering" refers to the practice of treating subsets or "tiers" of a provider network differently, usually in terms of the co-payments or coinsurance that patients pay out-of-pocket for seeing providers in a particular tier. For example, HPHC patients seeing specialists in tier one typically pay a \$15 copayment per visit, whereas patients seeing specialists in tier two pay \$25 per visit.

diminished somewhat, MMS still remains very concerned about GIC's tiering efforts and has questioned GIC's methodological approach.

- No physician groups that were interviewed reported making changes in their practice patterns as a result of GIC's data or their health plans' tiering decisions. However, many cited difficulty obtaining the types of detailed data that would enable them to increase their efficiency and move to a higher tier.
- GIC reported that it has not yet realized significant cost savings from network tiering, but acknowledges that co-payment differentials may be too small to incent beneficiaries to choose more efficient providers.
- Interviews with health plans and physicians suggest that future cost savings, as well as physicians' reactions to efficiency measurement, will likely depend on whether GIC increases the financial incentives that encourage beneficiaries to choose more efficient physicians. These incentives and the potential for market-wide savings will also increase if health plans begin selling tiered network products to other payers in the market.

BACKGROUND ON HEALTHCARE MARKET IN BOSTON

GIC, which self-administers health insurance for more than 285,000 employees and retirees of the Commonwealth of Massachusetts, is the largest single purchaser of health care services in the Boston area.¹⁰ To provide health services to its beneficiaries, GIC contracts with six of the seven largest health plans operating in Boston, including HPHC, Tufts Health Plan, and four smaller local plans. Notably, Blue Cross Blue Shield of Massachusetts—the area's largest health plan—does not contract with GIC.

In general, Boston is characterized by highly integrated physician groups and hospital systems. For example, the membership of Partners Healthcare (an integrated health care system that offers primary care and specialty services) includes Massachusetts General Hospital, Brigham and Women's Hospital, several community hospitals, several community health centers, and three major physician organizations totaling more than 4,500 physicians. Primary care physicians (PCP) are also commonly organized into large Independent Practice Associations (IPA). Although GIC is the largest single purchaser of health care, it is not a dominant player from the perspective of most IPAs, large multi-specialty groups, and integrated health care systems. In one large multi-specialty physician group with more than 700 physicians, for example, GIC beneficiaries represent less than five percent of their patient population.

¹⁰ The U.S. Census Bureau reports that the population of Massachusetts was 6,437,193 in 2006 and the Commonwealth Connector reports that 372,000 residents were uninsured according to a 2006 state survey. This indicates that GIC beneficiaries represent about 4.7 percent of all insured residents of Massachusetts.

GIC'S MEASURES OF PHYSICIAN RESOURCE USE

Motivations for Physician Resource Use Measurement

GIC is the primary driver of episode-based efficiency measurement of physician practices in the Boston area. GIC's efficiency measurement initiative—the Clinical Performance Improvement Initiative (CPI)—was implemented in 2004. GIC views efficiency measurement as an important tool for reducing the Commonwealth's health care costs in conjunction with its other cost containment efforts, such as raising beneficiary co-payments and tiering prescription drugs.

Some health plans, such as HPHC and Tufts Health Plan, performed episode-based efficiency analysis prior to the CPI, but had not used this information to create physician tiers. For example, in 2003, Tufts Health Plan began providing some specialists with confidential feedback reports that included episode-based results, and in 2004 it introduced the Navigator product, which created tiers for inpatient hospitals. The plans' primary motivation for increasing their use of episode-based efficiency measurement to profile physicians over the past several years has been to retain GIC beneficiaries by meeting GIC's tiering requirement. GIC's most recent contracts with health plans require the plans to submit claims data to GIC for their full books of business (not only for GIC members), which GIC then compiles into an all-payer dataset and analyzes using Symmetry Episode Treatment Groupers (ETGs). GIC also analyzes the same dataset to produce quality measures. Plans are required to use both types of measures to create tiered physician networks for GIC beneficiaries.

Methodological Considerations in Resource Use Measurement

Data Aggregation

GIC contracts with Mercer Inc. to aggregate data from all six GIC health plans and perform episode-based efficiency analyses. Data aggregation allows for the development of a single composite efficiency score that relies on a higher sample size for each physician than if calculated for individual plans. GIC considers 30 episodes of care per physician to be the minimum number necessary for accurate efficiency analysis; approximately two-thirds of the health plans' physicians have at least 30 episodes in the all-payer dataset in a given year. In several instances a physician would not have had sufficient episode volume with a single health plan to allow for a stable efficiency score calculation. In addition to efficiency score stability and reliability, data aggregation helps to provide a more complete picture of a physician's overall practice across plans. Without data aggregation, a substantial proportion of physicians would be characterized as more efficient than average based on claims data from one plan, but less efficient than average based on claims data from a different plan. For example, prior to data aggregation, about 21 percent of dermatologists appeared efficient in one plan, but inefficient in another plan in 2005.

Mercer Tool Selection and Modification

Mercer selected Symmetry ETGs because it already owned a license for the product and two of the six GIC health plans were familiar with the tool. Since beginning work for GIC in 2004,

Mercer has made several modifications to Symmetry ETGs in consultation with the health plans. These include the exclusion of some episodes that are not relevant to the specialties being profiled, as well as modification of the Symmetry ETG outlier logic to truncate rather than exclude outlier claims.

Three years of data are used for constructing episode-based efficiency scores¹¹; however, for state fiscal year (FY) 2008¹², Mercer weighted the most recent year of data more heavily than the first two. MMS suggested this modification, because it believes that more recent data are more representative of physicians' current practice patterns. From year to year, Mercer has observed some fluctuations in provider efficiency scores, but it cannot distinguish between fluctuations due to changes in the efficiency measurement methodology versus those due to changes in provider performance over time.

HPHC Tool Modifications

During 2006, the first year that tiering products were scheduled for introduction, HPHC performed its own episode-based analyses in order to address concerns that Mercer's results would not be available in time for HPHC to meet GIC's deadlines for constructing a tiered network. HPHC made some modifications to the basic Symmetry ETG software logic. These modifications included allowing an episode to be attributed to more than one physician, eliminating episodes that it did not feel were germane to each specialty HPHC had chosen to profile, and redefining the peer group for some physicians. For example, physicians who were not classified as pediatric specialists, but who treated a high percentage of children, were compared to other pediatric specialists in HPHC's analyses. Overall, HPHC found its results were generally consistent with those it received from GIC's all-payer dataset. For example, there was strong concurrence between the two sets of efficiency measures regarding the least efficient 15 percent of physicians in the plan.

USES OF EFFICIENCY MEASURES BY HEALTH PLANS

GIC requires all six of its health plans to use episode-based results to construct tiered physician networks. In creating such networks, the plans apply their own schedule of negotiated fees to the episode-based results to generate a plan-specific cost per episode and plan-specific physician efficiency measures. Plans have also incorporated quality measures in their tiered networks. The first tiered networks were launched in Spring 2006 and the first modifications to those networks were scheduled to take effect in July 2007. Since GIC has not been prescriptive about plans' tiering methodology, plans have developed several different models as discussed below. Because officials from HPHC were interviewed directly, the discussion below focuses on

¹¹ Mercer uses lagged data—for the FY 2008 analysis, for example, episode-based efficiency scores were calculated from 2003, 2004, and 2005 claims data.

¹² State FY 2008 began in July 2007.

HPHC's approach. GIC officials also provided an overview of the other five plans' approaches.¹³

Tiered Networks

Selection of Physicians to Profile

HPHC currently tiers nine physician specialties that account for 85 percent of their specialty evaluation and management (E&M) charges. HPHC decided to tier specialists because they observed a spread of 25 to 30 percent among specialists' efficiency scores, which suggested room for practice pattern changes. In contrast, efficiency scores only varied 8-10 percent among their PCPs. HPHC decided not to tier some specialties because it believes that these areas would be too politically challenging at this time (Oncology) or because the specialty was otherwise treated like primary care (OB-GYN).

The remaining five health plans have taken a variety of approaches to network tiering. Two plans tier all of their Massachusetts-based physicians, including all specialists and PCPs. Another plan tiers all of its PCPs, as well as five specialties. The remaining two plans tier selected specialties, but do not tier PCPs. Overall, GIC estimates that across the health plans, tiered network providers will account for 65 percent of GIC's medical expenditures in state FY 2008.

Group- vs. Individual-Level Tiering

GIC requires some element of individual physician tiering in the networks that went into effect in July 2007. HPHC currently tiers five of its nine specialties at the individual level, and the remaining physicians are tiered at the group level. The plan believes it is important to tier at the level at which physicians practice. That is, specialists that generally work as groups are better evaluated at the group level. Large IPAs are better evaluated at the small practice group level because decisions that affect efficient performance are not generally made at the IPA-level.

The remaining five health plans have taken a variety of approaches to group versus individual-level tiering. GIC estimates that the typical plan has tiered 15 to 20 percent of its specialist network at the individual level. Two plans make all tiering decisions at the individual level. The remaining three plans tier some physicians at the individual level and others at the group level.

¹³ The approaches used by the other five plans are discussed generally in the sections that follow. Appendix A outlines the details of each plan's approach based on information provided by GIC officials. HPHC was the only health plan that MedPAC and MPR staff interviewed directly.

Placement into Tiers

Physicians must first meet quality measurement goals, where measures of quality are calculated at the group level. The top 90 percent of physicians pass the quality threshold, as well as those who scored at least a 90 percent on the available measures, and those with quality denominators that are too small to calculate a reliable score. Next, physicians are evaluated on efficiency (at the group or individual level, depending on the specialty). On average, the most efficient 30 percent of specialists are selected for tier one. New physicians and those with too few episodes to be evaluated for efficiency are automatically placed into the second tier. From year to year, the tiers have been relatively stable, with movement of 10 to 15 percent of physicians between the tiers, mainly comprised of those who were just above and below the threshold to qualify for tier one the previous year.

The other plans use two main methods for physician tier classification. In one plan, physicians qualify for tier one through either their quality or efficiency performance, and new physicians are defaulted to tier one. The four remaining plans use both quality and efficiency measures to develop a single composite performance score, and designate physician tiers by applying cut-off points to the combined score. Similar to HPHC, one plan has chosen a cut-off point that places about 30 percent of physicians in tier one. The other plans' methodologies place 45 to 85 percent of physicians in tier one. HPHC avoided this approach because, in the context of some controversy about episode-based efficiency measurement, it believed physicians would find it less acceptable to single out poor performers than to recognize high performers.

Network Incentives

All plans use a \$10 co-payment differential between tiers to encourage beneficiaries to select the most efficient physicians. For example, HPHC patients seeing specialists in tier one typically pay a \$15 per visit copay, whereas patients seeing specialists in tier two pay \$25 per visit. GIC selected a small co-payment differential to "create a soft landing" for the new program, minimizing physician and beneficiary reactions while the tiered network products mature.

Confidential Feedback Reports

GIC is working with its physician advisory board to develop a template for physician feedback reports. At this time, only HPHC and Tufts Health Plan provide detailed physician reports. HPHC is also in the process of rolling out new efficiency reports, which will provide physicians with more detailed information than prior versions. HPHC's new reports indicate providers' overall efficiency scores, including their relative rankings for each major cost component (outpatient, inpatient, and pharmacy). The reports also include the five episode categories where a physician's costs diverge most from his/her peer group, and quantifies the resources that could be saved (or would be expended) if the physician's performance conformed to the peer group. Finally, the reports document performance for the top six highest volume episode categories for the physician's practice, also identifying resource use impacts.

Consumer Reporting

GIC and its plans have provided only high-level information to beneficiaries about tiering decisions. For example, through its online provider directory HPHC indicates which tier a physician has been assigned, but does not provide a rationale for tier designation (for example, by indicating that a provider is in tier two because of low episode volume). GIC notes in its consumer materials that plans assign physicians to tier one if they deliver high-quality and cost-effective care.

PHYSICIAN REACTION

Interactions with Physician Groups

Between GIC and Physician Groups

In the fall of 2003, before it had begun conducting episode-based analyses for the health plans, GIC convened a meeting with MMS to discuss its approach. MMS expressed some dissatisfaction with episode-based efficiency measurement, and in response GIC adopted several MMS recommendations for modifying the program. GIC also formed a physician advisory board, which is currently developing a model template for physician feedback reports and will advise GIC going forward.

Between HPHC and Physician Groups

HPHC has held town hall meetings with physicians to explain episode-based efficiency analysis and its tiering method, though attendance at these meetings has been somewhat less than HPHC anticipated. While HPHC has engaged some of its larger physician groups in detailed reviews of the efficiency data, overall feedback has been relatively low. Even when the plan shifted some physicians from tier one to tier two as part of the July 2007 network update, it received few inquiries from physicians.

Physician Reactions to Efficiency Measurement

Several physicians endorsed the general concept of efficiency measurement, but feel there are significant practical challenges that need to be overcome in order to operationalize it in an accurate and equitable manner. Risk-adjusting for patient population characteristics (e.g., health status, comorbidities, etc.), obtaining a sufficient sample size, selecting the appropriate episodes to profile for each specialty, and identifying the appropriate peer group of physicians for comparison are among physicians' chief concerns. Overall, physicians indicated that analysis at the group level is more appropriate than at the individual level for reasons of statistical validity and because physicians in the Boston area generally practice within medical groups.

Physician Reactions to Health Plans' Uses of Efficiency Measures

Although physicians were very concerned when GIC first announced its profiling efforts in 2004, GIC stated that its health plans have received fewer complaints in recent years. Physicians

confirmed the decline in their attention to this issue and attributed their response to the minimal impact GIC has actually had on their patient volume or finances to date. Some physicians suggested that the move to individual-level tiering would likely provoke a stronger response, since physicians are personally sensitive to being designated as tier-two.

While the overall physician response has diminished somewhat, MMS remains very concerned about GIC's tiering efforts and has questioned GIC's technical methodology. In November 2006, MMS commissioned a report which evaluated GIC's analytical approach and made 34 recommendations for improvement, some of which GIC has since adopted. MMS also stated that physicians face significant difficulties obtaining details on their efficiency scores from health plans, and it views the current approach as unfair to physicians who score poorly because they do not have enough information to contest the analyses. Given these issues, MMS is concerned about health plans selling their tiered network products to other purchasers in the market at this time.

Overall, physician concerns about GIC's use of efficiency data fall into three categories: concerns about different tiering designations across plans, concerns that data are not being used constructively, and concerns about beneficiaries' reactions. None of the six GIC plans use the same network tiering methodology, even though they use the same source data. Because of this variation, a physician may be placed into tier one for one plan, but into tier two for another. Physicians reported that they do not have sufficient time or incentives to analyze how each plan ranks their performance. Thus, although GIC has compiled an all-payer, complete book-of-business dataset, the process appears fragmented to physicians, and they do not understand how to change their behavior to perform more efficiently.

A related concern is that the data are not being applied constructively. Several physicians reported that they had not received underlying data from plans at a sufficiently-detailed level to permit them to take corrective action and improve their efficiency scores. Another consistent comment was that the data are too out-of-date for physicians to trust the scores as a credible reflection of their current practices. For example, data from 2003-2005 were used to develop July 2007 tiered networks. Several physicians commented that the underlying episode-based efficiency data could be more constructively applied by highlighting potential areas for improvement and ways to develop best practices, but that the data are not currently used for these purposes.

Finally, physicians are concerned about their patients' reactions to tiering decisions. Concerns run in both directions. Physicians are concerned that their patients will question why they were labeled tier-two physicians and will assume that the designation is due to poor quality. Conversely, physicians are also concerned about being labeled "cost-effective" tier-one physicians, since their patients might worry about receiving inadequate treatment.

Physician Uses of Efficiency Measurement

One physician group reported using episode grouper software independently for internal analysis; however, no physician groups that were interviewed reported making changes in their practice patterns as a result of GIC's data or the plans' tiering decisions. According to one group, GIC's financial incentives have not been sufficient to drive patient volume, and GIC beneficiaries generally represent only a small proportion of physicians' patient populations.

Several physician groups reported being receptive to re-examining their practice patterns, but indicated significant difficulties in obtaining data from the health plans that would allow them to identify areas for improvement. HPHC reported that it had discussed the data at a fairly detailed level with administrators from 10-12 physician groups who wanted to understand their efficiency scores, but that these groups represented only about 10 percent of the groups with whom they contract.

KEY LESSONS AND CONCLUSIONS

Effects of Efficiency Measurement on Resource Use

GIC reports that it has not realized significant cost savings through its application of episode-based efficiency scores and network tiers at this point, but acknowledges that co-payment differentials may be too small to incent beneficiaries to choose more efficient providers. HPHC has estimated savings of just 0.1 to 0.2 percent of premiums. If all patients moved from tier two to tier one, or all physicians in tier two matched the performance of their tier one peers, Mercer estimates a maximum savings potential of 7 to 8 percent of health care costs.

Key Benefits and Challenges

GIC's efforts in Boston illustrate some benefits and challenges of episode-based efficiency measurement. One key benefit is that GIC's all-payer dataset has laid the groundwork for developing accurate efficiency measures that can convey information to physicians about a large proportion of their practices. Also, allowing the six plans to design their own tiered networks has given GIC the flexibility to see which of several approaches will be best received by both the physician and beneficiary communities.

However, allowing plans to implement different tiering methods has also resulted in challenges. Diverse tiering approaches across the plans creates the impression of a fragmented system, diluting physician attention to the data. Realizing cost savings via the tiered network approach may require more coordination across health plans so that physicians perceive uniform incentives. Another challenge is that beneficiary reaction has been almost nonexistent. Strengthening the financial incentives for beneficiaries to select more efficient providers may be necessary to provoke a response from beneficiaries, which will in turn influence physician behavior. However, it is unclear whether the co-payment differentials necessary to produce significant cost savings would be politically acceptable.

Another major challenge is reporting data to physicians in a format that allows them to identify actions they can take to improve their efficiency scores. To date, GIC has hesitated to require health plans to perform these analyses because it believes physicians should take the initiative to analyze their own practice patterns. However, it is not clear that, even with sufficient financial incentives, most physicians have the level and type of resources necessary to perform that type of data analysis themselves.

Assessment of Future Market Trends

Future cost savings, as well as physicians' reactions to efficiency measurement, will likely depend on whether GIC increases the financial incentives that encourage beneficiaries to choose more efficient physicians. GIC also hopes to move towards better standardization of its plans' tiering methodologies and confidential feedback reports, as well as more individual-level tiering, all of which should strengthen physicians' incentives to use efficiency measurement data to adjust their practice patterns. In the short term, it appears that GIC will continue to be the primary motivator for efficiency measurement in the Boston area; however, several other purchasers have expressed interest in health plans' tiered network products. Accordingly, the potential for market-wide cost savings and the intensity of physicians' reactions will also depend on whether health plans begin to sell their tiered network products to other payers in the market.

TABLE 1
SITE VISIT SUMMARY HARVARD PILGRIM HEALTH CARE
BOSTON, MA

Aspects of Measuring Resource Use and Quality	HPHC (GIC) ^a
OVERVIEW	
Year began using physician resource use measures	2004
Episode grouper or other measurement product(s) used	Symmetry ETG
Quality measures used in conjunction with resource use measures	Yes, from Resolution Health Inc. (GIC's quality analysis contractor), HEDIS, AHRQ, and SAGE
TECHNICAL SPECIFICATIONS	
"Gated" approach vs. combination of quality, resource use measures, and/or sample size	Gated approach—1) quality, then 2) efficiency
Individual provider and/or group level measurement	Individual provider-level measurement, rolled up to group-level for tier designation for some specialties
Specialties profiled	9 overall 5 tiered at individual level (allergists, general surgeons, neurologists, ophthalmologists, otolaryngologists) 4 tiered at group level (cardiologists, dermatologists, gastroenterologists, orthopedic specialists)
Attribution method	Episode assigned to clinician with the highest clinician fees, as long as clinician is responsible for at least 25% of the episode fees charged. If no physician has at least 25 %, episode is unassigned.
Minimum number of episodes required for profiling a physician	30
Percentage of potentially eligible physicians profiled	65%
Addressing outliers	Truncation at the 2nd and 98th percentiles; exclusion of some low-volume episodes for some specialties
Methods for benchmarking/ranking	Quality—top 90% Efficiency—top 25-30% Comparison group is all Massachusetts physicians in the all-payer GIC dataset
Price standardization methods	GIC uses price-neutral analysis; HPHC applies contracted rates to GIC data
Data source and/or aggregation (including number of years of data)	Claims data from the full book of business for six GIC health plans; three years of data used for episode-based analysis with most recent year weighted more heavily
Key modifications over time	Modification of outlier methodology, exclusion of some episodes from some specialties, weighting of most recent year of data

Aspects of Measuring Resource Use and Quality	HPHC (GIC) ^a
USE OF RESOURCE USE MEASURES	
Tiering	Yes, two-tiered Harvard Independence Plus Plan with \$10 beneficiary co-payment differential
Physician feedback	Yes, confidential reporting of cost areas driving episode-based efficiency score, and financial impact of outlier episode categories
Pay-for-performance	No
Consumer reporting	Yes, web-based portal where beneficiaries can view tiering designation
Key changes over time	Tiering physicians at the individual instead of at the group level

^a Some methodological decisions were made by GIC rather than by HPHC.

TABLE 2
SUMMARY OF OTHER GIC PLAN TIERED NETWORK PRODUCTS

	Specialties Profiled	Tiering Methodology	Percent of Physicians in Tier One
Tufts Health Plan	Cardiologists tiered at group level 7 specialties tiered at individual level (dermatology, endocrinology, gastroenterology, neurology, ophthalmology, orthopedic specialists, otolaryngology)	Cut-off point applied to composite efficiency and quality score	30%
Unicare	All specialists and PCPs tiered at the individual level	Qualify for tier one through either efficiency or quality score	45%
Health New England	5 specialties tiered at individual level (cardiology, dermatology, gastroenterology, orthopedic specialists, otolaryngology) Internal medicine and family practice (excluding pediatricians) tiered at group level (into three tiers)	Cut-off point applied to composite efficiency and quality score	80% specialists PCPs: 10% tier one 80% tier two 10% tier three
Neighborhood Health Plan	All PCPs tiered at group level 3 specialties tiered at individual level (cardiology, endocrinology, OB-GYN)	Cut-off point applied to composite efficiency and quality score	85%
Fallon Community Health Plan	All specialists and PCPs tiered at group level Six types of physicians eligible for tiering at the individual level if in a tier-two practice (internal medicine, family practice, pediatrics cardiology, endocrinology, gastroenterology,)	Cut-off point applied to composite efficiency and quality score Physicians in certain tier-two practices evaluated individually for placement in tier one	85%

APPENDIX D

AUSTIN SITE SUMMARY

Contract No.: RFP03-06-MedPAC/E4016631
MPR Reference No.: 6355-300

MATHEMATICA
Policy Research, Inc.

**Site Visit Summary:
Austin Texas**

August 1, 2007

*Stephanie Peterson
Timothy Lake*

Submitted to:

Medicare Payment Advisory Commission
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Project Officer:
Niall Brennan

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Timothy Lake

CONTENTS

	Page
SUMMARY	D.1
BLUE CROSS BLUE SHIELD OF TEXAS' MEASURES OF PHYSICIAN RESOURCE USE	D.2
Motivations for Physician Resource Use Measurement	D.2
Selection of Physicians for Resource Use Measurement.....	D.3
Methodological Considerations in Resource Use Measurement.....	D.3
HEALTH PLAN USE OF RESOURCE UTILIZATION MEASURES	D.4
Inclusion in Network.....	D.4
Confidential Feedback Reports	D.5
Consumer Reports	D.5
Approaches of Other Austin-Area Health Plans	D.6
PHYSICIAN REACTION	D.6
Physician Uses of Resource utilization Measurement	D.6
Interactions between Blue Cross Blue Shield and Physician Groups	D.6
Physician Reactions to Resource Use Measurement.....	D.7
Physician Reactions to Health Plans' Uses of Resource Utilization Measures	D.7
KEY LESSONS AND CONCLUSIONS.....	D.8
Effects of Measurement on Resource Use	D.8
Key Benefits and Challenges	D.9
Assessment of Future Market Trends	D.9

SITE VISIT SUMMARY AUSTIN, TEXAS

This report summarizes information obtained in a series of interviews with health plan officials and physicians conducted during a site visit to Austin, Texas, from June 26-27, 2007, to explore health plans' use of episode groupers for resource utilization measurement. Following a semi-structured interview guide, MedPAC and Mathematica Policy Research, Inc. (MPR) staff spoke with executives from Blue Cross Blue Shield of Texas (BCBSTX), the Texas Association of Family Physicians (TAFP), and two area physician group practices—one a large multi-specialty clinic and the other a medium-sized internal medicine group with four physicians. One of the physicians in the latter group was also a member of the Texas Medical Association. The purpose of the interviews was to learn about Blue Cross Blue Shield's technical experiences using the Symmetry episode treatment grouper (ETG) and later, the Medstat episode grouper (MEG) to measure physician resource utilization, as well as to obtain information about physicians' reactions to private health plans' use of this type of measurement.

SUMMARY

- BCBSTX, the primary health plan in the Austin area, views performance measurement as a tool to address rising costs of health care by providing employer groups with a lower cost healthcare option that maintains evidence based measure performance for their employees.
- Since 2003, BCBSTX has used episode grouper software (Symmetry ETG for two years and currently Medstat MEG) to identify a subset of physicians and professional providers within their BlueChoice network. The Risk Adjusted Cost Index (RACI) which reflects differences in the allowed costs of episodes of care after adjustment for case-mix, severity of illness, comorbidity of patients, physician specialty, and geographic region is used to define the subset. This subset is called the BlueChoice Solutions network, and serves as an alternative lower-premium insurance product offered to employers.
- Beginning in 2006, BCBSTX also started using evidence based measures calculated by Health Benchmark, Inc. (HBI) software as another prerequisite to inclusion in its BlueChoice Solutions network.
- Also beginning in 2007, BCBSTX launched a web-based transparency tool available to the general public that rates PPO participating physicians and professional providers on evidence based measure performance and affordability.
- BCBSTX has engaged physicians in its program. It mails detailed reports providing resource utilization data at the Tax-ID level to physicians and offers a review process where physicians or groups can appeal their designations. BCBSTX has also recently formed a committee to review future evidence based measures with members appointed by the Texas Medical Association. Local physician groups are represented on the committee.

- In general, the physicians we spoke with in the Austin area believe that health plans should investigate inappropriate resource utilization as long as the data the health plans use are made available to physicians for review.
- Austin area physician group concerns over the BCBSTX measurement program fall into two broad categories: 1) the validity of the data (including concerns about using claims-based data) and 2) communication and reporting methods (both to physicians and the public).
- The physician groups feel that providing data at the individual provider level is the most effective way for individual physicians to improve their care; nevertheless, some physician groups have started trying to make resource utilization and evidence based measure performance improvements based on the group level data they receive from BCBSTX.

BLUE CROSS BLUE SHIELD OF TEXAS' MEASURES OF PHYSICIAN RESOURCE USE

Motivations for Physician Resource Use Measurement

BCBSTX began using episode grouping software in 2003 as a way to address rising health care costs. Employer groups had been requesting cost-effective alternatives to the existing provider networks (including the BlueChoice network). In response, BCBSTX has developed physician resource utilization measures by analyzing claims data with Symmetry ETG and Thomson MEG software tools. In conjunction with evidence based measures, these data are used to identify a subset of physicians and professional providers within the current BlueChoice network. The subset—the BlueChoice Solutions network—is a high performance network alternative for employer groups seeking to provide lower cost healthcare options to their employees. BCBSTX also markets this product as a lower cost alternative to individuals. The subset network is offered as an alternative to the broader BlueChoice network under Preferred Provider Organization (PPO) benefit plans, and Point-of-Service (POS) benefit designs.

BCBSTX also recently developed a web-based transparency tool on affordability and performance on evidence based measures that the general public can view. Affordability is displayed as a 5-point rating scale, which reflects differences in the costs of care a BCBSTX physician (or group) manages compared to other BCBSTX physicians (or groups) within the same specialty treating similar types of member patients in same geographic area. Performance on evidence based measures is displayed as one of two different shaded blue ribbons or no ribbon at all based on an aggregated evidence-based measures. A physician (based on the performance of his/her same specialty colleagues within a practice aggregate score¹⁴) receives a blue ribbon if his/her group is recognized for commendable (light blue ribbon) or outstanding performance (dark blue ribbon) compared to same specialty peers in the BCBSTX network. If a certain threshold was not reached or there are no measures for the specialty or there was insufficient data no ribbon is displayed.

¹⁴ If a physician practices in more than one group (meaning s/he practices under more than one Tax-ID), s/he can receive more than one ribbon designation.

Selection of Physicians for Resource Use Measurement

Most physician specialties that are ambulatory-based are included in the selection process (including primary care). Hospital-based specialties, such as radiology, anesthesiology, neonatal, pathology, and pediatric subspecialties, are excluded from the resource utilization measurement process.

Methodological Considerations in Resource Use Measurement

BCBSTX has periodically refined its resource utilization measurement methods since it began to conduct resource use measurement in the early 2000s. BCBSTX first used Symmetry ETGs, which were selected because the company had experience using that software product for group reporting purposes. It used this software for the first two iterations of conducting resource use measurement, and found that as physicians reviewed the data challenges were made to the clinical relevance of the groupings and adjustment for severity. When the contract with the software company ended, BCBSTX explored the market for grouper software, compared what was available, and decided to switch to Medstat MEG software.

BCBSTX is satisfied with the Medstat MEG software. One of the reasons Medstat MEG is attractive is the extent of severity and risk adjustment it provides. Output from the software is used to calculate a Risk-Adjusted Cost Index (RACI), which is a ratio of the total *allowed* costs of qualified episodes for which the provider is attributed responsibility divided by the total *expected* costs. The expected cost is based on the average cost of qualified episodes partitioned by MEG, severity of illness (which is classified on an ordinal scale from 0 to 3), comorbidity of patients (defined by the DxCG Relative Risk Score, RRS), provider specialty, and geographic region. The RACI measures the extent to which episodes of care managed by a physician or professional provider are more or less costly than peers within the same specialty in the same geographic region after adjustment for case-mix, severity of illness and comorbid conditions of the patients.

As part of its resource utilization measurement process using MEGs, BCBSTX has made a number of refinements over time with respect to the data included in episodes in the standard software. For instance, BCBSTX does some additional trimming of the actual claims data after the episodes are constructed and before expected costs and RACI's are calculated. It has excluded such items as high-cost drugs and air ambulances from the measurements. For example, BCBSTX found that the use of an air ambulance could be an issue for rural doctors and in some cases substantially increased their episode costs. In addition, there are episodes associated with certain services, such as respiratory syncytial virus (RSV) injections, that BCBSTX does not want to discourage so they are also removed from episode cost calculations. In the MEG 430:Preventive Services, eye care episodes have been refined significantly by BCBSTX. Eye care services and medical services are segregated to prevent one from confounding the other. Another refinement by BCBSTX, is splitting infant care episode categories into more discrete categories by age to address differences in recommended care for infants with febrile illness.

Other refinements BCBSTX has made include examining the total number of episodes in the state attributed to various physician specialties for the entire year, to assess whether certain specialties should be held accountable for certain categories of episodes. BCBSTX's refined the

trimming parameters related to scope of practice in response to complaints by some physicians about inclusion of conditions for which they did not have primary management responsibility. If there are fewer than 50 episodes of a clinical condition statewide within a given year attributed to physicians in a particular specialty (for example, fewer than 50 cases of OB-GYN treating a flu episode), episodes in that Medical Episode Group (MEG) will not be included in calculating the RACI of anyone in that specialty in that year. Only episodes in a specialty's scope of episode categories—based on the statewide volume of episodes included within a category—will be attributed to the physician.

In early 2006, BCBSTX also began using evidence-based indicators driven by HBI software to exclude physicians with the lowest evidence based measure scores from the BlueChoice Solutions network. (Evidence based measures are calculated using both claims and enrollment data.) BCBSTX contracted with HBI for 36 evidence based measures applicable to 25 specialties (allergy & immunology, neurology, OB-GYN, oncology, cardio specialties, general surgery, family practice, among others) that it currently includes in its program, including: preventive screening, diabetes, childhood immunizations, antibiotic use, medication adherence, mental health/other, medication monitoring, respiratory illness, ophthalmology, heart disease, cancer/other, musculoskeletal, and avoidance of complications.

HEALTH PLAN USE OF RESOURCE UTILIZATION MEASURES

BCBSTX uses resource utilization measures to produce confidential physician feedback reports, for public reporting to consumers, and to create an alternate network with lower cost providers without lower performance on evidence based measures (the BlueChoice Solutions network).

Inclusion in Network

Construction of the Network

In order to qualify for the BlueChoice Solutions network, physicians must meet various prerequisites. The first prerequisite is that the physician/professional provider must already be part of the BlueChoice network (because BlueChoice Solutions is a subset of the BlueChoice network). BlueChoice credentialing criteria include: provider qualifications to practice in a specialty, actions of licensing boards, actions of medical staff committee on clinical privileges, member complaints, malpractice cases, and Medicare/Medicaid sanctions.

These physicians must then meet additional BlueChoice Solutions credentialing criteria, which include having their RACI score below a defined threshold and having a minimum of 30 qualified episodes. Beginning in 2007, physicians must also be above a threshold level on an aggregate score evidence based measure to be included in the BlueChoice Solutions network, as described above. Inclusion in the network is re-evaluated annually.

Network Incentives

In general, the BCBSTX product steers members towards BlueChoice Solutions physicians/professional providers. There is no direct financial reward for physicians/professional providers who participate in the BlueChoice Solutions network. ,

BCBSTX has also launched a web-based transparency tool that the general public can view, which might influence consumer choice of providers in the larger BlueChoice network. The tool provides information on physicians' rankings on affordability (based on the RACI score) and performance on evidence based measures (based on an aggregate of the measures applicable to their specialty).

Confidential Feedback Reports

BCBSTX provides reports at the Tax-ID level. A Tax ID level report can reflect the experience of a single physician in solo practice or several hundred physicians in a large group. In response to physician requests, the number and structure of reports has expanded over time. Routine reports mailed to provider groups provide information on the RACI, the MEGs, severity, comorbidity group, total allowed cost, and total expected cost. Additional reports made available on request provide information on Comorbidity Index relative to peer group, component cost and utilization compared to peer group, and procedure code content compared to peer group.

With the inauguration of the transparency component on Provider Finder, Affordability Scale and EBM display information was mailed to 40,000 network physicians. The EBM display consisted of either the dark or light blue ribbon if the group reached a particular threshold compared to their peers. After consultation with physician organizations, no indicator was displayed for those with low performance or insufficient data or no measures for the specialty. More detailed reports were made available that included the percentile of the aggregated evidence-based measure score as well as detail for each indicator including the number of physicians in the group measured for that indicator, the number in the denominator (the count of patients who qualified for the indicator) and numerator (the count of patients who were provided services satisfying the criteria for the indicator) as well as both the provider and the BCBSTX specialty rate for each indicator.

Consumer Reports

There are two main ways consumers can be affected by the results of BCBSTX analyses: 1) through their list of in-network physicians each year, and 2) through the web-based-information available online. First, because BCBSTX invites physicians to the BlueChoice Solutions network based on their resource utilization and evidence based measure scores, consumers may no longer have access for in-network benefits to those providers who fall out of the Solutions network. They can view which providers are in the Solutions network online through BCBSTX's "Provider Finder."

Secondly, BCBSTX launched a web-based reporting tool in April 2007 that allows consumers to make comparisons among providers based on affordability and evidence base measure data for more than 40,000 physicians. As described above, evidence based measure

performance is indicated by a light blue, dark blue, or no ribbon and physician affordability is measured on a 5-point sliding scale. This tool is embedded in its online “Provider Finder,” but is available to anyone (whether or not a BCBSTX member).

Approaches of Other Austin-Area Health Plans

BCBSTX is the primary player in the Austin area. The other two major health plans are United and Aetna, although they have much smaller market shares and currently do not appear to be using the ETG or MEG efficiency-based analysis in the area.

PHYSICIAN REACTION

Physician reaction to resource utilization measurement is mixed. Some physician groups are using BCBSTX resource use analysis to identify opportunities for resource utilization or evidence based measure performance improvement. Others believe that the BCBSTX resource use is flawed. Overall, physicians’ concerns about the use of resource use measurement fall into two broad categories. Physicians have concerns about 1) the validity of the data (including concerns over using claims-based data), and 2) communication and reporting methods (both to physicians and to the public).

Physician Uses of Resource utilization Measurement

Some physician groups in Austin are using BCBSTX resource use measures in an attempt to change practice patterns and lower their RACI or to increase evidence based measure performance. For example, some physicians stated they now try to discuss costs of different procedures with their patients, such as MRIs, before providing such services. Another medical group noted that it did not score well on the screening for Chlamydia measure. Before it received its BCBSTX report, this internal medicine practice did not realize that it should be including chlamydia screening—when performing Pap smears. In response to the report, it now includes this type of screening.

BCBSTX noted that a number of the physician groups that have responded to its analysis (for example, by requesting more data) had lower RACIs and wanted to figure out ways to lower it further or maintain their position relative to peers in order to remain in the network.

While some medical groups in the BCBSTX Austin area network have been using the MEG data, others do not use it or respond to their RACI. Several physician groups stated that BCBSTX needs to provide more information on how its preferred network provider designations are determined before they will take such measurement seriously. Several stated that the data BCBSTX is currently using are flawed (but did not expound on this) and are consequently not taking the measurements seriously.

Interactions between Blue Cross Blue Shield and Physician Groups

BCBSTX has tried to be proactive in engaging physicians in its BCBSTX program. Before and after the program went live, BCBSTX made a series of presentations on about methodology

to the Texas Medical Association's Council on Socioeconomics, various county medical societies, medical group practices, and independent physician associations.

Once the BlueCompare transparency program began, however, physicians with concerns became more vocal. In response, BCBSTX formed a formal committee on measures of performance (for both resource use and evidence based performance improvement) with members appointed by the Texas Medical Association. In addition, in response to the Texas Medical Association's concerns over gray ribbons in the BCBSTX web-based transparency tool, BCBSTX postponed the public launch of it. Before taking the website live to the public four months later, BCBSTX deleted the gray ribbon, which represented physicians who either did not meet the evidence based measurement criteria or did not have sufficient data or were in a specialty with no measures. The Texas Medical Association believed that the gray ribbon was confusing for consumers to understand because the color grouped together both poor performing providers and providers who BCBSTX did not have enough information about.

BCBSTX offers processes for review of transparency program displays and network eligibility on physician request. Reviews are available for RACI and evidence based medicine score. The largest volume of review requests in the current year involve the evidence based measure score, which drives the ribbon designation. BCBSTX stated this is probably because it is a very new program.

Physician Reactions to Resource Use Measurement

Physicians' reactions to the concept of resource use measurement varied in the Austin area. Some physician groups have responded positively to BCBSTX analyses, requesting more data on how to lower their RACI, improve evidence based measure performance, and/or stay in the Solutions network. One medical group stated the process is an opportunity for physicians to learn and provide better care. Other medical groups responded with more skepticism to measuring and holding physicians accountable for resource use, stressing that patients should have more of a role in cutting costs. For example, several physicians stated that for liability reasons it was difficult to not give a patient an MRI if that patient requested one, even if the physician believed it was not necessary. Some physicians also stressed that there are unintended consequences of using resource use measurement, such as doctors pushing high cost or noncompliant patients out of their practices because they might lead to a lower designation. In turn, physicians stated that these patients would turn up at the emergency room, which would be more costly overall for the system.

Physician Reactions to Health Plans' Uses of Resource Utilization Measures

In general, the physicians we spoke with in the Austin area believe that inappropriate resource utilization merits analysis by health plans as long as the underlying data the health plans are using is made available to the physicians. However, most believe that BCBSTX analysis needs improvement before they will take it seriously. Physicians' concerns fall into two broad categories: 1) validity of data (including concerns about the use of claims-based data) and 2) communication and reporting methods (both to physicians and to the public). First, many physician groups are concerned about the use of claims-based data to calculate resource utilization scores. One physician stated that he compared his electronic medical records (EMRs)

to the report sent to him by BCBSTX and the BCBSTX data did not match his records. Another physician had a problem with claims-based data because he felt it was too delayed, stating that data can be a year or a year-and-a-half old, which often leads to inappropriate results since diagnoses and guidelines change frequently. A third physician cited an AMA study which found that claims are under-coded for, or are not coded for, 30 percent of the time. Some medical groups acknowledge that EMRs are not a magic solution to documentation either and may not be feasible or affordable for all practices to implement.

Some physicians are more concerned that BCBSTX data are not clean enough to be used for resource utilization measurement. One group stated that it received a report that included physicians who were not even part of the group practice. Another physician said he got “dinged” for not performing liver screening that he had performed, but it was not captured in the report because it was included in a more comprehensive set of tests he had done that was not recognized by BCBSTX. The physicians we spoke with all agree on the importance of health plans providing underlying resource utilization data at the individual physician level. However, they do not trust the validity of data that they cannot view in its entirety.

A second concern over BCBSTX’s use of resource utilization and evidence based measures deals with communication and reporting methods - both to physicians and to the general public. The physicians we spoke with all stressed that more communication and education are needed because current reports are difficult to understand. Others stressed that under current reporting, there is no credit given for incremental improvements and that BCBSTX uses arbitrary benchmarks/thresholds to compare physician resource utilization or evidence based measures against. Some thought improvement was difficult given that BCBSTX only provides reports at the medical group level. One physician stated it is hard to motivate the lower-performing providers in the group without individual-level information to show the specific differences between group members. In addition, many have concerns over the BCBSTX transparency web-based tool that is available to the public. This tool categorizes physician groups with different colored ribbons depending on how they rate on the evidence based measures. Physicians do not like this tool because not all specialties are included. They feel that if customers view this web page and do not see their doctors with blue ribbons (even if it is because that specialty is not measured, or there is just not enough data), it will hurt their reputations.

KEY LESSONS AND CONCLUSIONS

Effects of Measurement on Resource Use

The medical groups we spoke with remain concerned about resource use measurement, although some of these groups have begun to implement changes - for resource utilization as well as evidence based measure performance improvement - based on the reports and designations they have received from BCBSTX. However, all agreed that more communication, education, and input from physicians are needed before physicians will really take such programs seriously.

Key Benefits and Challenges

BCBSTX stated that it has seen changes in physician behavior as a result of resource utilization measurement. In addition, several physicians we spoke with asserted they have tried to become more proactive with patients about whether to conduct certain procedures that they may not need (such as MRIs), by discussing potential costs with them. In addition, some physicians have implemented practice changes based on their evidence based measure score.

However, large challenges still remain with using the data to change behavior. Providers stated they need more easily understandable data to make additional improvements. All the physicians we spoke with also stressed the importance of having access to individual-level data in order to incent lower-performing physicians to improve and to increase their perception of the data's validity.

Assessment of Future Market Trends

One physician we spoke with said that physicians tend to go through several stages that are similar to the stages of grief (for example, denial, anger, and then acceptance) before they finally accept resource utilization and evidence based measure improvement changes. Most of the physicians we spoke to seem to have left the denial stage, accepting that these measurements are part of the future of health care, especially given BCBSTX's dominance in the Austin area. However, some of the larger practices are less dependent on BCBSTX designations and still dismiss what BCBSTX is doing as a short-term marketing pitch to employers. It therefore remains to be seen whether or not these tools can influence physician behavior with respect to resource use (or evidence based medicine) over the longer term.

SITE VISIT SUMMARY TABLE
BLUE CROSS BLUE SHIELD TEXAS (BCBSTX)
AUSTIN, TEXAS

Aspects of Measuring Resource Use and Evidence Based Medicine	BCBSTX
OVERVIEW	
Year began using physician resource use measures	2003
Episode grouper or other measurement product(s) used	Symmetry ETG (first two years). Currently, Medstat MEG (since 2004)
Evidence Based Medicine measures used in conjunction with resource use measures	Yes, effective January 1, 2006, evidence-based measures included as criteria for inclusion in Solutions network. These measures come from Health Benchmarks, Inc. (HBI). 25 specialties and a total of 36 Evidence Based Medicine measures.
TECHNICAL SPECIFICATIONS	
“Gated” approach vs. combination of Evidence Based Medicine, resource use measures, and/or sample size	Combination approach: 1) resource use measures (RACI score), and 2 evidence-based measures (EBM Score).
Individual provider and/or group level measurement	Group-level (profile on the Tax-ID level)
Specialties profiled	Most specialties with some very specific specialty exceptions (such as neonatologists and hospital based providers)
Attribution method	<p>The provider who bills the greatest total Relative Value Units (RVUs) for a given episode.</p> <p>When there is no provider specifically identified by RVUs, the episode is attributed to the provider billing the greatest number of outpatient evaluation or management services for the episode.</p> <p>When no provider is identified by either of the above, the episode is attributed to the provider with the highest allowable cost included in the episode.</p>
Minimum number of episodes required for profiling a physician	30 qualified episodes
Percentage of potentially eligible physicians profiled	72-74%
Addressing outliers	<p>Limit period of analysis to 2 years</p> <p>Claim period is 30 days for episode to be considered complete, although sometimes this varies by condition</p> <p>Limit incomplete episodes (if fewer than 30 episodes, must have some type of fee minimum)</p> <p>A minimum amount of beneficiary membership is required for the patient to be included in the analysis and this varies by condition</p> <p>Exclude claims outside the state</p> <p>Also exclude a small number of episodes (such as elective abortions; behavioral health care by PCPs)</p> <p>If a specialty treats something fewer than 50 times throughout the entire state, the type of episode is not included in resource utilization calculations</p>

Aspects of Measuring Resource Use and Evidence Based Medicine	BCBSTX
Methods for benchmarking/ranking	Resource utilization and Evidence Based Medicine measurements are based on thresholds
Price standardization methods	No price standardization
Data source and/or aggregation (including number of years of data)	Two years
Key modifications over time	1) Changed from Symmetry ETG to Medstat MEG in 2004; 2) Added evidence based measures 1/1/2007; 3) Added transparency website piece in 4/2007
USE OF RESOURCE USE MEASURES	
Tiering	BlueChoice network is one of the prerequisites to be included in the BlueChoice Solution network, the subset of physicians designated by BCBSTX as more affordable
Physician feedback	Yes, group level reports issued (at Tax-ID level)
Pay-for-performance	No
Consumer reporting	Yes, web-based portal where anyone can view BCBSTX designation on affordability (RACI score) and evidence based measure performance (EBM Indicators)
Key changes over time	Considering establishing a 3-tier product (BlueChoice providers, normal PPO, out of network providers)

APPENDIX E
CLEVELAND SITE VISIT

Contract No.: RFP03-06-MedPAC/E4016631
MPR Reference No.: 6355-300

MATHEMATICA
Policy Research, Inc.

**Site Visit Summary:
Cleveland, Ohio**

August 16, 2007

*Stephanie Peterson
Timothy Lake*

Submitted to:

Medicare Payment Advisory Commission
601 New Jersey Avenue, NW
Suite 9000
Washington, DC 20001

Project Officer:
Niall Brennan

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
Timothy Lake

CONTENTS

	Page
SUMMARY	E.1
BACKGROUND ON HEALTHCARE MARKET IN CLEVELAND	E.2
UNITEDHEALTHCARE’S MEASURES OF PHYSICIAN RESOURCE USE	E.2
Motivations for Physician Resource Use Measurement	E.2
Selection of Physicians for Resource Use Measurement	E.3
Methodological Considerations in Resource Use Measurement	E.3
HEALTH PLAN USE OF EFFICIENCY MEASURES	E.4
Reporting Methods	E.4
Approaches of Other Ohio-Area Health Plans	E.5
PHYSICIAN REACTION	E.6
Physician Uses of Efficiency and Quality Measurement	E.6
Physician Reactions to Efficiency Measurement	E.6
Interactions between UnitedHealthcare and Physician Groups	E.7
Physician Reactions to Health Plans’ Uses of Efficiency Measures	E.7
KEY LESSONS AND CONCLUSIONS	E.8
Effects of Efficiency Measurement on Resource Use	E.8
Key Benefits and Challenges	E.8
Assessment of Future Market Trends	E.9

SITE VISIT SUMMARY CLEVELAND, OHIO

This report summarizes information obtained in a series of interviews with health plan officials, physicians, and medical society representatives conducted during a site visit to Cleveland, Ohio, from July 18-19, 2007, to explore health plans' use of episode groupers for efficiency measurement. Using a semi-structured interview guide, MedPAC and Mathematica Policy Research, Inc. (MPR) staff spoke with executives from UnitedHealthcare, the Academy of Medicine of Cleveland and Northern Ohio (AMCNO), the Ohio State Medical Association (OSMA), and physician and management representatives from two large health care systems in the area: the Cleveland Clinic and the University Hospitals System. The purpose of the interviews was to learn about UnitedHealthcare's technical experiences using episode treatment grouper (ETG) and anchor target procedure grouper (ATPG) software to measure physician efficiency, as well as to obtain information about physicians' reactions to private health plans' use of this type of measurement.

SUMMARY

- In response to studies showing the need for quality improvement in healthcare in the United States,¹⁵ UnitedHealthcare began measuring physician quality and resource use in Cleveland, Ohio in 2005. UnitedHealthcare implemented a program called the UnitedHealth Premium Designation program, based on two previous UnitedHealthcare programs. It evaluates both quality and efficiency use and markets the product to employers as a "quality first" evaluation tool.
- Physicians who pass a quality threshold, determined either by national disease and specialty standards or by UnitedHealthcare's Scientific Advisory Boards, receive a star designation. Only those physicians deemed high quality in this manner proceed to an efficiency use analysis. Physicians who are determined to also be cost-efficient receive a second star.
- There are two major health systems in the Cleveland area-the Cleveland Clinic Health System and the University Hospital Health System. Most large physician group practices are associated with one of these two systems and have had little direct experience or knowledge of resource-use measurement by health plans. This is most likely because such information is handled at a higher level than individual provider, such as by the health systems quality management staff.
- In general, the majority of physicians remain skeptical about resource use measurement. While most physicians are amenable to health plans' efforts to

¹⁵ See 1) Elizabeth McGlynn et al. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine*. Vol 348(26): 2635-2645. June 26, 2003, and 2) Institute of Medicine. *Crossing the Quality Chasm: A New Health Care System for the 21st Century*. March 2001.

measure quality of physician care, some have concerns about quality measurement as well.

- Many physicians prefer electronic medical records (EMRs) to claims data for measuring quality and resource use. Many also believe that quality measures should incorporate indicators of consumer satisfaction. All physicians agreed on the importance of standardization for resource use and quality measurement.
- Few physicians have started using UnitedHealthcare's reports to implement changes in resource use; however, some have started implementing quality improvement efforts based on both internal quality programs and health plan resource use and quality reports.
- Physicians agreed that improvement of the products used to measure efficiency and quality require more communication with and education from the health plan.

BACKGROUND ON HEALTHCARE MARKET IN CLEVELAND

The major health plans in the Cleveland market include UnitedHealthcare along with Medical Mutual of Ohio and Anthem Blue Cross Blue Shield. In particular, UnitedHealth Group provides insurance coverage for 1,176,521 million (47,78000 in HMO; 941,002 in POS/EPO; 135,208 in PPO; and 52,531 in the Medicare product) employees and retirees in Ohio.

In general, however, Cleveland is characterized by two major health systems—the Cleveland Clinic Health System and the University Hospitals Health System. The Cleveland Clinic and the University Hospitals Health systems are the two major players dominating most of the hospital market. These two systems are also among the area's largest employers—and this includes employing most of the large physician group practices in Cleveland. Other large employers in the area include General Electric, which has begun steering employees to certain providers based on UnitedHealthcare's Premium designation program

UNITEDHEALTHCARE'S MEASURES OF PHYSICIAN RESOURCE USE

Motivations for Physician Resource Use Measurement

In response to studies showing variation in care and the need for quality improvement in health care in the United States, UnitedHealthcare implemented a program in Cleveland, Ohio in 2005 called the UnitedHealth Premium Designation program. UnitedHealthcare markets the program as a "quality first" evaluation tool. The health plan denotes network physicians who meet or exceed the quality benchmarks with one star. Physicians who are designated for quality proceed to an efficiency of care analysis; those who meet the efficiency of care benchmark get a second blue star.

The program goal is to help consumers make more informed choices in selecting physicians and hospitals, as well as to help individual physicians increase quality and cost efficiency. For the Premium Designation program, UnitedHealthcare chose Symmetry ETG, Symmetry EBM Connect and Anchor Target Procedure Grouper (ATPG) software to measure resource use and quality. UnitedHealthcare staff indicated that other grouper software could have been used for

measuring resource use, but the plan chose software already owned by UnitedHealthcare's parent company, UnitedHealth Group.

Selection of Physicians for Resource Use Measurement

Physicians must be in UnitedHealthcare's network to qualify for its Premium Designation program and are eligible if they are in one of 16 specialties with established quality standards (based on national disease and specialty standards or metrics determined by UnitedHealthcare's Scientific Advisory Boards (SABs)). These physician specialties include: 1) allergy, 2) cardiothoracic surgery, 3) cardiology, 4) endocrinology, 5) family medicine, 6) infectious disease, 7) internal medicine, 8) pediatrics, 9) nephrology, 10) neurology, 11) oncology,¹⁶ 12) orthopedic surgery, 13) pulmonology, 14) rheumatology, 15) obstetrics-gynecology, and 16) neurosurgery (spine).

In addition, the physician must be board-certified in his/her specialty as determined by the American Board of Medical Specialties or the American Osteopathic Association. These physicians are divided into either cognitive/non-procedural specialties or procedural specialties, and are first evaluated on quality. If they pass the quality threshold, they are then evaluated for efficiency. Premium star designations are determined annually.

Fifteen of the specialties in UnitedHealthcare's network are automatically evaluated for designations. The exception, oncologists, must apply annually for the quality designation by completing an application.

Methodological Considerations in Resource Use Measurement

UnitedHealthcare uses Symmetry episode treatment grouper (ETG), Symmetry EBM Connect and anchor target procedure grouper (ATPG) software to evaluate quality and resource use measurement.¹⁷ Quality measurement is done first and differs somewhat for two subcategories of specialties: proceduralists and non-proceduralists (or cognitive specialists). For example, UnitedHealthcare uses Symmetry's EBM Connect tool and rules are based on national standards for constructing quality indicators for cognitive or non-proceduralists (allergists, OB/GYNs, pediatricians, and internists, among others). These standards are set by national consensus organizations, such as the National Quality Forum (NQF) and the AQA Alliance. UnitedHealthcare also prioritizes conditions referenced in the 2001 Institute of Medicine study on quality improvement (see reference in footnote 1) and AQA by weighting these conditions. Where EBM rules do not exist, UnitedHealthcare has established external SABs made up of leading specialists to advise on rules and standards. SABs have been established for the following specialties: musculoskeletal care, cardiac care, and cancer care.

¹⁶ Oncologists are reviewed on quality but not efficiency use because claims-based data used to measure efficiency do not include information on cancer disease stage.

¹⁷ UnitedHealthcare stated that conceptually ATPG and ETG software is similar and they could have probably used ETG for both analyses. However, the ATPG is different from ETG in that it is not illness-based but procedurally based and that surgeons tend to resonate with it more than ETG software.

Quality is measured at a national level, i.e., the quality standard is the same in all markets. In order to receive a quality star designation, cognitive or non-proceduralists must receive a score of at least 70 points out of a total of 100 possible points. These physicians must also have at least a sample size of five patients for the evaluation. For proceduralists, ATPG is used and physicians' scores that fall within an accepted confidence interval are designated with a quality star. In addition, primary care physicians and endocrinologists are able to receive credit toward their quality star designation if they participate in the NCQA Diabetes Physician Recognition program or the NCQA Heart/Stroke Recognition program.

Only physicians who meet quality benchmarks proceed to the efficiency of care analysis. All specialties are measured at the market level and each specialty is compared only to other physicians in their specialty. Resource use is based on risk-adjusted completed episodes. Outlier physician efficiency scores are also adjusted using a modified Windsor technique, where episodes that are below the 5th percentile of resource use are discarded and episodes above the 95th percentile are truncated to the 95th percentile. Seventy-five percent of a typical physician's UnitedHealthcare specific practice is assessed.

Efficiency of care for cognitive or non-proceduralists is measured again using Symmetry ETG software, and for proceduralists it is measured using ATPG software. For proceduralists, for example, resource use measurement includes professional fees, diagnostic testing, inpatient facility costs, and follow-up ancillary monitoring testing. A 95 percent confidence interval is constructed around each physician's data and is used to determine premium designation.

Although physicians are first analyzed at the individual level, UnitedHealthcare also uses a sequential methodology to evaluate group practices. The health plan aggregates the data on all physicians by specialty in a given medical group practice and then runs the above processes, treating each specialty within a group as an individual physician. Group methodology can only be used to help individual physicians who either have insufficient data to be designated on their own or meet quality on their own but not efficiency. For example, if a physician is not given a second star on his/her own, but their specialty in the medical group is given an efficiency star, then the physician also receives a second star.

HEALTH PLAN USE OF EFFICIENCY MEASURES

UnitedHealthcare uses efficiency measures for public reporting to consumers, confidential physician feedback reports, and its pay-for-performance initiative.

Reporting Methods

Public Reporting

During the study period, UnitedHealthcare product offerings generally did not use financial incentives (such as lower co-payments) to steer enrollees to Premium providers. Instead, the health plan attempts to steer enrollees toward physicians with a Premium designation via display of designation information on the consumer portal and in some cases through employer communication to their employees. The program is designed to give consumers and employers the ability to make more informed choices in selecting their physicians and hospitals. United

Healthcare reports to each large customer annually on the percentage of cost and numbers of employees treated by Premium designated physicians during the prior year measurement period. Employers such as General Electric have begun to use the data to construct reports in order to help educate employees on quality and use of Premium designated doctors.

UnitedHealthcare maintains a consumer web site that displays Premium star designations for its network providers. It is available to UnitedHealthcare both customers as well as the general public online at the myuhc.com web site, an online provider directory. Premium designation is shown at the individual practicing physician level.

Confidential Feedback Reports

UnitedHealthcare mails confidential feedback reports to physicians eligible for the Premium program 30-45 days prior to the public display of designation results. The reports include web site and login information that allows physicians to view more detailed reports electronically. The electronic reports (which can be 60 to 70 pages long) are downloadable and contain patient level detail. The goal of the first report is to give physicians time to correct UnitedHealthcare's analysis prior to posting results on the consumer web site if they have issues with their designation. UnitedHealthcare provides a process where the physician can present his/her argument for reconsideration; UnitedHealthcare says it accepts physician self-reported data as valid, —although it reserves the right to audit the physician if it so chooses. Physicians who have insufficient data display the text “insufficient data with UnitedHealthcare” next to their name. Physicians in a specialty that is not evaluated display the text “specialty not evaluated” next to their name. The website explains that a lack of designation may occur for many reasons as listed above but does not identify doctors specifically who did not pass quality. UnitedHealthcare stated that the majority of physician reconsideration requests relate to capture of testing data (for example, a beneficiary filling a drug prescription at a VA hospital instead of at a registered UnitedHealthcare pharmacy would not be captured).

Pay-for-Performance

UnitedHealthcare has started a limited pilot pay-for-performance program (in Cleveland, Ohio, and Chicago, Illinois) based on its quality and efficiency measures (as well as a third dimension of administrative efficiency). Top performing physicians are rewarded with a percentage increase in their fees. The pay –for –performance program is rolling out in additional markets in 2007.

Approaches of Other Ohio-Area Health Plans

UnitedHealthcare is one of three major health plans in Ohio and is the most advanced in its use of efficiency-based analyses. The Medical Mutual of Ohio (MMO) and Anthem Blue Cross and Blue Shield (BCBS) are the other two major health plans. The MMO has a pilot transparency program that operates in the Canton, Ohio, area; Anthem BCBS released a pilot transparency tool program in Dayton, Ohio, in 2006 and in May 2007 expanded the program to Cincinnati (as well as other areas outside Ohio, including Lexington and Louisville, Kentucky). In particular, the Anthem BCBS program maintains a consumer web site that displays cost

information based on claims data for nearly 40 different medical procedures. Anthem's goal is to provide consumers with more choice and to steer them to more cost-effective health care. Anthem's program does not include a quality measurement component.

PHYSICIAN REACTION

Physician reaction to efficiency measurement remains somewhat skeptical. Physicians have concerns over the validity of using claims data to measure resource use. While most physicians are more amenable to health plans' efforts to measure quality, some physicians have concerns over quality measurement as well. All the physicians we spoke with agreed on the importance of standardization among efficiency use and quality measurement tools. In addition, physicians agreed that more dialogue is needed between health plans and physicians in order to improve products used for measuring quality and efficiency.

Physician Uses of Efficiency and Quality Measurement

Some physicians have started to use efficiency and/or quality measurement results to alter their practice behavior. One of the large health care systems we spoke with has its own internal quality improvement programs. One of these programs is designed for 250 of their primary care physicians. The hospital system sends performance reports, developed from their EMR data, to physicians each quarter detailing results on a variety of preventive measures, including diabetes and screenings for breast, cervical, and colon cancer. The hospital system also is involved in quality improvement programs with the Robert Wood Johnson Foundation.

Some physicians in smaller practices have also started to try to improve quality. However, in some cases, for these physicians, the drive to improve quality has been based on reports received from UnitedHealthcare. For example, one physician said he was not aware that he was not providing a specific test to diabetes patients as often as he should, so he implemented a system of reminders to increase the number of his patients receiving the test.

Few physicians, though, have instituted resource use changes based on UnitedHealthcare reports, and none could cite specific examples of such practice modifications. Most of the physicians were unfamiliar with the reports, most likely because they were part of a large practice that did not filter the results down to the individual level. For example, one physician we spoke with stated that he used to receive reports from health plans when he was in a solo practice, but since joining one of the areas large health systems, he no longer receives such reports. He stated that the health system has in place a quality department that deals with such paperwork directly. However, UnitedHealthcare stated it has begun to see changes in referral patterns to physicians with Premium star designations.

Physician Reactions to Efficiency Measurement

Most physicians stated they would respond to quality and efficiency use reports because they want to do the right thing and improve their quality of care. Most physicians as well as the management staff and the large health systems also stated that they would pay attention to reports coming from Medicare that captured a significant percent of their caseload. However, a

few physicians involved with one of the two larger healthcare systems in Cleveland stated that reports comparing them to other physicians would not necessarily be useful because often physicians at hospitals are so specialized they might be the only person in the hospital performing certain tests. In addition, some stated they are too busy to review their data, especially given that they did not think the program offered much benefit to their practice.

The physicians we spoke with also remain skeptical of the validity of using claims-based data to measure resource use. Some physicians stated that the use of EMR data is preferable. However, many of the physicians not associated with the two main systems in Cleveland (the Cleveland Clinic or the University Hospital System) worry about spending an enormous amount of money on EMR software only to find out later that it is not compatible with requirements from the Center for Medicare and Medicaid Services (CMS) or health plans. All physicians stressed the importance of standardization for both resource use and quality measurement. In addition, some physicians also stressed the importance of patient accountability when measuring resource use and quality.

Interactions between UnitedHealthcare and Physician Groups

UnitedHealthcare started an outreach campaign six to eight months before the program began to try to “socialize” physicians. Part of the campaign included UnitedHealthcare staff traveling to different areas of the country to speak with physicians about the UnitedHealth Premium Designation program. UnitedHealthcare’s program in Ohio is based in part on feedback from physicians from the plan’s prior quality and resource use measurement programs, including one in St. Louis, Missouri. Results of these dialogues included adding the quality-first criteria piece, and establishing a reconsideration process. UnitedHealthcare now mails physicians the results of the Premium program’s quality and efficiency use analysis prior to publicly posting the star designations to give physicians time to review those designations and request reconsideration before consumers can view them on the web site.

However, some physicians stated that better communication is needed between them and UnitedHealthcare. Some physicians said they had tried to contact UnitedHealthcare to receive more information about their designation but response was slow. Some physicians stated it did not seem like UnitedHealthcare was prepared to provide them with more information, which is probably why it took so long to respond; and even when the plan did respond, it was unable to provide them with the information they needed. One of the larger healthcare systems in the area also requested aggregated data from UnitedHealthcare, but the plan has only been able to provide it with individual level data, a problem for the affiliated hospital system where hospitals share patients among many physicians.

Physician Reactions to Health Plans’ Uses of Efficiency Measures

Physicians at one of the two larger systems in the area were less aware of UnitedHealthcare’s designation program than those in smaller practices because the larger systems have dedicated quality improvement staff to collect the information. However, representatives from this hospital system said they are still trying to figure out the Premium designation program and therefore has not taken an official stand on the program. However, this quality improvement team did state that UnitedHealthcare’s use of quality and efficiency

measurement needs improvement. They also said that more dialogue is needed between physicians and health plan officials. In addition, they said physicians are often unclear about how to use their reports to improve practices. One physician said he received a report saying he was inefficient. When he spoke with the health plan to get more detail, the plan told him he had a low sample size, which the physician stated is not a solution for improvement.

While most physicians we spoke with agreed that quality measurement is more important than efficiency measurement, they also had concerns about quality measurement. One of the larger systems, in particular, has issues with the fact that some of their international world-renowned physicians did not get a star quality designation because they lacked board certification. The group stated that it is not sure if consumers care about the designations, but it did not want something like this to damage the group's reputation. Other physicians are concerned about UnitedHealthcare's methodology for measuring quality. Some do not feel that claims data can measure quality effectively and that it can be misleading to consumers, who are more likely to think of quality in terms of outcome measures and customer satisfaction.

KEY LESSONS AND CONCLUSIONS

The medical groups we spoke with remain skeptical about quality and resource use measurement. Physicians have concerns with the validity of the data, especially when claims data is used for measuring both efficiency and quality. Physicians asserted that EMRs should be used to capture both efficiency and quality. Some also stated that consumer satisfaction is an important measurement of quality as well. All physicians stressed the importance of standardizing such programs and that Medicare should take the lead in doing so.

Effects of Efficiency Measurement on Resource Use

UnitedHealthcare's program is still in its infancy, although the plan stated that it is starting to see shifts in referral patterns toward designated physicians and that employers have started implementing benefit packages based on the designations. Few physicians we spoke with, however, have implemented resource use changes as a result of UnitedHealthcare's reports, although some have started implementing quality improvement programs. Physicians seem to be more likely to react to the results of the data if they do not receive the quality or efficiency designation.

Physicians also stated that more information and communication is needed from health plans before they are likely to take such programs seriously, and that standardization of measurement among all the different health plans and Medicare is necessary as well.

Key Benefits and Challenges

UnitedHealthcare has seen a few changes in physician referral patterns; but since the program is still new, it is too early to determine this definitively. Some employers have also started designing employee benefit packages based on the designations.

Physician education and “socialization” remain a challenge. Many physicians are still skeptical about the data and are not sure how the programs will affect them. One medical society representative reported that many of its members have little understanding of the programs and do not have time to pay attention to such initiatives. In addition, many physicians stated they do not think this is something consumers care about; and, in some cases, such as with the larger healthcare systems in the area, consumers do not have a choice in the specialist they see at the hospital anyway.

Assessment of Future Market Trends

UnitedHealthcare plans to continue marketing their Premium Designation program in the Cleveland market as well as other areas nationally. Although they acknowledge that there is a huge learning curve for physician participation, they state that they have begun to see physician referral patterns shift as a result of the program. However, the program is still in its infancy so UnitedHealthcare also stated that a more data is still needed to determine this directly. In addition, the physicians we spoke with see efficiency use and quality measurement as part of their future but most are still trying to understand how it will affect them. All physicians agreed that the more that such measures can be standardized, the more likely physicians will become invested in the process. Future cost savings, as well as physician reactions to them, will likely depend on the health plan’s ability to communicate effectively and convey the relevance of the program to the physicians and two major health care systems in the area.

SITE VISIT SUMMARY TABLE
UNITEDHEALTHCARE
CLEVELAND, OHIO

Aspects of Measuring Resource Use and Quality	UnitedHealthcare
OVERVIEW	
Year began using physician resource use measures	August 2005 (however, it is an outgrowth and evolution of two of the health plan's previous performance programs, which began in 2003).
Episode grouper or other measurement product(s) used	Symmetry EBM Connect, Symmetry ETG and Anchor Target Procedure Grouper (ATPG)
Quality measures used in conjunction with resource use measures	Yes, only physicians that receive a star for quality are analyzed for efficiency.
TECHNICAL SPECIFICATIONS	
"Gated" approach vs. combination of quality, resource use measures, and/or sample size	Gated approach in the sense that physicians must pass quality threshold before they proceed to an efficiency measurement.
Individual provider and/or group level measurement	Both. Group methodology may override the individual designation if higher, as long as individual provider did not fail. (Group is based on the combined results of the individual specialists within the group, not an average).
Specialties profiled	Primary care physicians and most specialties are included. (16 included: allergy, cardio-thoracic surgery, cardiology, endocrinology, family medicine, infectious disease, internal medicine, pediatrics, nephrology, neurology, oncology, orthopedic surgery, pulmonology, rheumatology, OB-GYN, neurosurgery)
Attribution method	For cognitive or nonproceduralists: based on majority of claims dollars spent during an episode For proceduralists: based on proceduralist who submitted the claim for the interventional procedure
Minimum number of episodes required for profiling a physician	20 episodes (for procedural) 10 episodes (for cognitive)
Percentage of potentially eligible physicians profiled	Around 69%. (Roughly 58% are designated as either 1 star or 2 star; 11% do not meet quality criteria.; and there is insufficient data for 31%).
Addressing outliers	There is outlier logic built into the efficiency of care analysis to address both low and high cost outliers. (Exclude low cost cases that fall below the 5 th percentile for the market for that specific episode. Include high cost cases but truncate the patients' episode costs at the 95 th percentile for the market cost for the specific episode).

Aspects of Measuring Resource Use and Quality	UnitedHealthcare
Methods for benchmarking/ranking	For the efficiency measure, confidence intervals determine the threshold as to whether designated efficient or not. Quality is also based on a pre-determined threshold
Price standardization methods	No price standardization
Data source and/or aggregation (including number of years of data)	Two years of claims data
Key modifications over time	Additional specialties included; new conditions and rules used (approximately 100 new rules added); standardization methodology introduced (i.e., confidence intervals; outlier methodology; minimum # of cases increased; Q scores); standardized terminology; scorecard enhancements; expansion of the breadth of conditions evaluated for quality
USE OF RESOURCE USE MEASURES	
Tiering	No
Physician feedback	Yes, reports are mailed to physicians at individual and group level. Physicians receive log-in information to access detailed data online (at patient level) as well.
Pay-for-performance	Yes, a piloted automatic fee schedule enhancement is built into the system, which takes into account both quality and efficiency scores, as well as an added third dimension of administrative
Consumer reporting	Yes, the consumer web site, myuhc.com, shows the star designations. (The web site is an online physician directory).
Key changes over time	Adopted confidence interval methodology